Comparative Genomics of Toll Like Receptors: Investigating the impact of Selective Pressures on TLR Evolution & Divergence

Project report submitted to Sister Nivedita University in partial fulfilment of the requirement for the degree of

MASTER OF SCIENCE
IN
BIOTECHNOLOGY

Submitted By
Meghana Rani Dalei
Registration ID: 230020219450
Enrollment ID: 2322210002021

Advisor
Dr.Manojit Bhattacharya
Assistant Professor
Department of Zoology
FAKIR MOHAN UNIVERSITY

School of Health Science and Translational Research
Sister Nivedita University
Kolkata
June 2025

Comparative Genomics of Toll Like Receptors: Investigating the impact of selective pressures on TLR evolution & divergence

Project report submitted to Sister Nivedita University in partial fulfilment of the requirement for the degree of

MASTER OF SCIENCE
IN
BIOTECHNOLOGY

Submitted By
Meghana Rani Dalei
Registration ID: 230020219450
Enrolment ID: 2322210002021

Advisor
Dr. Manojit Bhattacharya
Assistant Professor
Department of Zoology
Fakir Mohan University

School of Health Science and Translational Research
Sister Nivedita University
Kolkata
June 2025

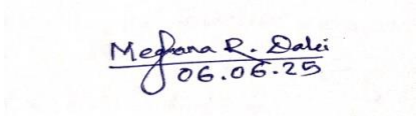
DECLARATION BY THE STUDENT

I hereby declare that the work which is being presented in this dissertation entitled "**Comparative genomics of Toll Like Receptors: Investigating the impact of selective pressure on TLR Evolution and divergence**" in partial fulfilment of the requirements for the award of the degree of Masters in Biotechnology is an authentic record of my own work carried out during the period from 1st January 2025 to 30th April 2025 under the supervision of Dr. Manojit Bhattacharya, P.G. Department of Zoology, Fakir Mohan University, Balasore, Odisha.

I acknowledge that, I have read and understood the UGC regulations, 2018 (*Promotion of Academic integrity and Prevention of Plagiarism in Higher Educational Institutions*).

I confirm that, this dissertation has been checked for plagiarism using the plagiarism checking software "Turnitin". At the end of the dissertation, a copy of the summary report demonstrating similarities in content and its potential source (if any), generated online using plagiarism checking software is enclosed. I herewith confirm that, the dissertation has been less than 10% similarity according to the plagiarism checking software's report and meets the UGC Regulations as well as plagiarism rules of Sister Nivedita University.

I further declare that, no portion of the dissertation or its data will be published without the Institute's or supervisor's permission. I have not previously applied for any other degree or award using the topics and findings described in my dissertations.



(Signature of the students with Date)

Name of the Student- Meghana Rani Dalei
Registration no.- 230020219450
Department of - Biotechnology

SISTER NIVEDITA UNIVERSITY
সিস্টার নিবেদিতা বিশ্ববিদ্যালয়
DG Block - 1/2, Action Area I, New Town, Kolkata - 700156, West Bengal, India.

DEPARTMENT OF BIOTECHNOLOGY & DEPARTMENT OF MICROBIOLOGY
জৈব প্রযুক্তি বিভাগ ও অণুজীব বিদ্যা বিভাগ

TO WHOM IT MAY CONCERN

This is to certify that the project report titled "**Comparative genomics of Toll Like** Receptor**: Investigating the impact of selective pressure on TLR evolution and divergence**", submitted by Ms.**Meghana Rani Dalei** bearing Registration No-**230020219450** and Enrolment No-**2322210002021** of the Department of Biotechnology, School of Health Science and Translational Research at Sister Nivedita University, Kolkata, is a bonafide record of research conducted by her for the partial fulfilment of the requirements for the Master of Science degree.

Date: 03.06.2025

Place: SISTER NIVEDITA UNIVERSITY

Signature of Head of the Department

Seal

Head
Department of Biotechnology
Sister Nivedita University
Block DG-1/2, Action Area 1
New Town, Kolkata-700 156

Acknowledgement

I take this opportunity to express my deepest gratitude to all those who have supported and guided me throughout the course of this research work. First and foremost, I would like to express my sincere thanks to my supervisor, **Dr.Manojit Bhattacharya**, for their invaluable guidance, continuous encouragement, and constant support throughout the duration of my dissertation. Their expertise, insightful suggestions, and patience were crucial to the successful completion of this work.

I extend my heartfelt thanks to **Prof. Bhaskar Behera,** Head/Coordinator and all the faculty members of the P.G. Department of Zoology, Fakir Mohan University, Balasore, Odisha for allow me and providing the necessary academic environment, resources, and technical support during my research. I am grateful to the laboratory and office staff, PhD Scholars and Project staff for their cooperation and assistance during my project work. I sincerely thank my friends, and seniors for their constant encouragement, constructive feedback, and for creating a positive and motivating atmosphere throughout my academic journey. I owe my deepest gratitude to Sister Nivedita University faculty members, Dr.Rajat Pal(Assistant Director) and Dr. Sudipta Saha (Assistant Professor) for their advice, encouragement, and help over these few months. Finally, I thank the Almighty for giving me the strength, health, and wisdom to successfully complete this dissertation.

LIST OF FIGURE Page no

1. Fig1: Flow chart showing the overall methodology used . 22
2. Fig 2: Bar graph showing of amino acids variation of TLR protein across species. 27
3. Fig2.2: Bar graph showing of instability index variation of TLR protein across species 27
4. Fig2.3: Bar graph showing of Aliphatic index variation of TLR protein across species 28
5. Fig2.4 :Bar graph showing of molecular weight variation of TLR protein across species 28
6. Fig2.5: Bar graph showing of Estimated half life variation of TLR protein across species 29
7. Fig2.6: Bar graph showing of GRAVY variation of TLR protein across species.(29)
8. Fig2.7: For species relation between theoretical pi and estimated half life. (30)
9. Fig 2.8: for TLR types molecular weight variation across species.(31)
10. Fig 3 : Mol wt and extinction coefficient appear to form a cluster with 1 outlier.(32)
11. Fig4 : This graph represents Aliphatic index and extinction coefficient having weak negative correlation suggests little to no direct relationship. (32)
12. fig 5: This scatter plot shows the eigenvalues (Yaxis) against the Principal components (Xaxis). (33)
13. Fig 6. This shows a Correlation loading plot of all physicochemical parameters. (34)
14. Fig7 : Correlation plot showing the variation in all physicochemical properties TLR protein across different species.(34)
15. Fig 8: It represents Principal component of analysis (**PCA**) scatter plot generate using PAST SOFTWARE, illustrating the distribution of various TLRs based on their physicochemical properties.(36)
16. Fig 9: Simplified cladogram showing phylogenetic relationship of different TLRs shows the evolution of vertebrates. (37)
17. Fig 10: Multiple sequence alignment of TLR proteins. (38)
18. Fig 11 : This is a circular phylogenetic tree (radial cladogaram). It visually represents evolutionary relationships among different Toll like receptors. (40)
19. **LIST OF TABLE :** TABLE 1: Descriptive statistical analysis of physicochemical properties of TLR protein across species. **26**

ABSTRACT

Toll Like Receptors (TLRs) are vital component of innate immunity system responsible for detecting pathogen associated molecular patterns (PAMPs) and initiating host defense responses. This study performs a comparative genomic and structural analysis of TLRs across five vertebrate classes-humans, frogs, birds, fish, and reptiles to investigate the impact of selective pressure on TLR evolution and divergence. The structural and physicochemical properties of TLR proteins were evaluated to understand their stability, hydropathicity, and functional domains. Multiple sequence alignment was conducted to identify conserved motifs. Using Bioinformatics tools, we performed multiple sequence alignments, Phylogenetic analysis and structural modeling to identify conserved and variable regions with in TLRs focusing on the Leucine-Rich Repeat (LRR) and Toll interleukin-1 receptors (TIR) domains. Our findings reveal distinct patterns of divergence in the extracellular LRR domain, correlating with ligand specificity and host pathogen interactions. Additionally sequence variations and structure adaptations in the TIR domain suggest evolutionary pressure shaping signaling efficiency .This study also explores the association between TLR variation and susceptibility to infectious and autoimmune disease in humans. By integrating sequence and structural data with disease associated Polymorphism studies, we provide insights into how evolutionary charges in TLRS influence immune function and disease risk. These findings contribute to a deeper understandings of TLRs evolution and pave theory for targeted interventions of immune related disease.

Key words :

TLR(Toll like Receptors), Phylogenetic Analysis, natural selection, comparative genomics, evolutionary divergence , selective pressure, positive selection, Physicochemical properties, Principal component Analysis, Innate immunity.

Abbreviations :

1. **NCBI :** National centre of Biotechnology Information
2. **PDB :** Protein Data Bank
3. **TLR :** Toll Like Receptor
4. **LRR :** Leucine-Rich Repeats
5. **TIR :** Toll/ interleukin-1 Receptor
6. **PAMP :** Pathogen Associated molecular Pattern
7. **MSA :** Multiple Sequence Allignment
8. **PCA:** Principal Component Analysis
9. **GRAVY:** Grand Average of Hydropathicity
10. **MW** : Molecular Weight

11. **Pi :** Isoelectric point

12.Dn/Ds : Nonsynonymous /Synonymous substitution Ratio

13.**DNA :** Deoxyribonucleic Acid

14. **Aa :** Amino acid

15.PAST : Paleontological statistics software

16.3D: Three dimensional

17.**MEGA :** Multiple Evolutionary genetic Analysis

18.PYMOL : Python Molecular graphic tool

19.RCSB : Research Collaboratory for structural Bioinformatics

20.FASTA : Fast – All (Text based format for sequence data)

CONTENTS

1. INTRODUCTION 12

2. OBJECTIVES 16

3. REVIEW OF LITRATURE 19

4. MATERIALS AND METHODS 23

5. RESULTS AND DISCUSSION 44

6. CONCLUSION 45

7. APPENDIX 46

8. REFERENCES 52

INTRODUCTION

The innate immune system serves as the first line of defense against pathogens and relies heavily on pattern recognition receptors (PRRs) to detect conserved molecular signatures known as pathogen-associated molecular patterns(PAMPs). Among the various PRRs, **Toll-Like Receptors (TLRs)** are one of the most extensively studied due to their central role in innate immunity and inflammation (Akira et al., 2006).

TLRs are type I transmembrane proteins characterized by an extracellular leucine-rich repeat (LRR) domain for ligand recognition and a cytoplasmic Toll/interleukin-1 receptor (TIR) domain responsible for downstream signaling (Takeda & Akira, 2004). In vertebrates, TLRs recognize a broad range of microbial components, including lipopolysaccharides (TLR4), double-stranded RNA (TLR3), and unmethylated CpG DNA (TLR9)**,** allowing them to initiate specific immune responses tailored to the nature of the pathogen (Kawai & Akira, 2010).

The diversity and distribution of TLRs across species reflect the evolutionary pressure exerted by microbial environments. Gene duplication, divergence, and loss have contributed to lineage-specific expansions or reductions in TLR repertoires, especially in non-mammalian vertebrates like amphibians, birds, and fish (Roach et al., 2005; Quiniou et al., 2013). Comparative genomics provides a powerful approach to understanding how TLR evolution is shaped by selective forces**,** particularly positive selection, which often targets regions involved in ligand binding and specificity (Wlasiuk & Nachman, 2010).

Structural evolution of TLRs is not merely a reflection of sequence divergence, but also affects ligand recognition and downstream signaling fidelity**.** Subtle alterations in the LRR domain may expand or restrict ligand binding, while changes in the TIR domain may influence adaptor protein interactions (Bell et al., 2003). The combined use of sequence alignment, phylogenetics, and protein structure modeling enables a comprehensive understanding of how molecular evolution drives functional diversification in the TLR family**.**

This study aims to investigate the evolutionary trajectory and structural implications of TLR diversification across representative vertebrates—humans, birds, amphibians, reptiles, and fish**.** B**y** analyzing patterns of selective pressure and mapping them onto 3D structural models**,** we seek to elucidate how evolutionary changes have impacted TLR function, stability, and ligand recognition**.**

The immune system is a complex network of cells, Tissues, and molecules. The ability of the immune system to recognize molecules that are broadly shared by pathogens is in part due to the presence of immune receptors called Toll-Like Receptors (TLRs). TLR named after the Toll protein first found in the fruit fly *Drosophila melanogaster* is a group of evolutionary ancient membrane-bound sensors localized to plasma membrane and endosomes (Anderson et al.1985). They are expressed on the membrane of leukocytes including dendritic cells, macrophages, and natural killer cells. It is the ancient form of host defense against pathogens in multicellular organisms, which functions as the first line of defense against foreign pathogens and invaders. There are many different types of TLRs ranging from 10 in humans to at least 17 in amphibians. Each TLR is unique with different properties that allow it to recognize distinct ligands. TLRs are expressed in many cell types of tissue and they recognize ligands of bacteria, fungi, and other microorganisms capable of transmitting diseases. TLRs were first identified in humans in the 1990s are membranes of the type transmembrane receptors family and evolutionary conserved protein among vertebrates and invertebrates (Medzhitovetal., Nature(1997). Adaptations to the complete pathogenic environment is a major aspect of the ecological adaptations of vertebrates, imposing frequent evolutionary pressure on molecules that form the direct interface between the host and pathogen because of the link between TLR dysfunction& several diseases in humans. The evolution of TLRs across vertebrates, including mammals, birds, reptiles, fish, and amphibians reflects a complex interplay of gene gain/loss, positive selection, and purifying selection(Voogdt & Putten.,2016). TLRs are ancient receptors with origins tracing back to early metazoans. Single- cysteine cluster TLRs, sharing architectural similarities with vertebrate TLRs, predate the divergence of bilaterians and cnidarians (Liu et al., 2020). The emergence of vertebrate TLRs occurred shortly after vertebrate origins, with 26 TLRs identified across vertebrates, including six (TLR3, TLR7, TLR8, TLR14, TLR21, and TLR22) present before the Agnatha-Gnathostomata split (Khan et al., 2019). Phylogenetic studies reveal eight TLR

subfamilies, with TLR3 as the most ancient, predating deuterostome divergence (Liu et al., 2020). Early TLR evolution involved dynamic gene gain/loss, leading to lineage-specific diversification. For example, birds lost TLR8 and TLR9 but gained unique TLRs like TLR21, while mammals acquired TLR6 and TLR10 (Khan et al., 2019) (Temperley et al .,2008).

Recent advances in computational biology have facilitated the in-depth analysis of protein structure, sequence alignment, and evolutionary patterns. Tools such as MEGA for Phylogenetic inference, PAST for statistical analysis, and PYMOL for 3D structure visualization have enabled researchers to explore the molecular and functional properties of TLRs in great detail. Furthermore, dN/dS ratio analysis provides insights into the strength and direction of selective pressures acting on protein-coding genes, offering a quantitative approach to studying molecular evolution (Yang & Bielawski, 2000).

The evolution of TLRs is linked to their functional roles in pathogen recognition, with specific codons under positive selection indicating adaptations for recognizing diverse pathogens(Kumar, 2016). Variations in TLR structure, such as the number of Leucine-rich repeats, suggest species-specific adaptations that enhance immune responses to local pathogen challenges (Ishengoma & Agaba, 2017). While the evolutionary patterns of TLRs indicate a strong influence of selective pressures, it is also important to consider that some TLRs may exhibit conservation due to their fundamental roles in immune function, suggesting a balance between adaptation and preservation in immune system evolution. In this study, a comparative approach was employed to analyze the TLR gene family in humans, frogs, birds, reptiles, and fish. The goal was to investigate how selective pressures have shaped the evolution, sequence variability, structural divergence, and physiochemical properties of TLRs. By integrating sequence alignment, phylogenetic analysis, physiochemical profiling, and structural visualization, this research aims to shed light on the adaptive evolution of innate immune receptors across vertebrate taxa.

The significance of this work lies in its contribution to understanding the molecular mechanisms under pinning immune system evolution. Since TLRs are among the ancient and conserved components of the immune response, studying their diversity across vertebrates not only

enhances our knowledge of host-pathogen evolution but also has practical implications for vaccine development, immunotherapy, and comparative immunology.

This work is important because it helps us understand the molecular mechanisms behind the evolution of the immune system. Given that TLRs are some of the oldest and most conserved elements of the immune response, examining their variations in vertebrates advances our understanding of host-pathogen co-evolution and has applications in immunotherapy, comparative immunology, and vaccine development (Bailey et al., 2014; Schenk et al., 2009).

Thus, this thesis aims to address key questions: How have TLRs evolved across different vertebrate lineages? What are the structural and functional implications of sequence divergence? And how do environmental and evolutionary pressures shape the molecular characteristics of these crucial immune receptors? Through an integrative genomic and structural analysis, the study aspires to uncover the evolutionary footprints left by pathogens on the TLR family, and to better understand the molecular basis of innate immune adaptation.

OBJECTIVES

To Investigate the ancestor of human TLR genes by comparing them with those of other vertebrates and non-vertebrates. Construct phylogenetic trees to identify divergence points. The research aims to perform comparative genomic analysis of vertebrate Toll-like Receptors (TLRs) across various species including mammals, birds, reptiles, amphibians, and fish, to understand evolutionary history of human TLRs and their divergence from other species (vertebrates).

Sequence Analysis: Use genomic data (from PDB, NCBI) to study TLR gene sequences across populations. Focus on conserved regions, mutations, and functional domains.

Population Genetics: Study the variability of TLR genes in human populations. Correlate genetic diversity with susceptibility to diseases and autoimmune disorders.

Comparative genomics: Compare human TLRs with those in primates or other mammals. Highlight differences in pathogen recognition and immune responses.

REVIEW OF LITRATURE

1. Overview of TLR: An important part of the innate immune system, Toll Like Receptors (TLRs) are crucial for identifying pathogen-associated molecular patterns (PAMPs). They were first identified as a component of embryonic development in *Drosophila melanogaster*, but their immunological importance was later revealed when it was discovered that mammalian homologs induced immune responses (Anderson et al.1985). Extracellular Leucine-Rich Repeats (LRRs), which are responsible for pathogen detection, and an internal Toll/Interleukin-1 receptor (TIR) domain, which starts signaling chains that result in inflammatory reactions, are characteristics of TLRs, which are membrane-spanning proteins. When TLRs are activated, cytokines are produced and co stimulatory molecules are up regulated, both of which are necessary for the start of adaptive immunity.

2. Evolutionary Perspective of TLRs Across Species – Diversification and conservation among vertebrates are both evident in the evolutionary history of TLRs. Although the core set TLRs expressed by vertebrate genomes is generally conserved, there can be lineage-specific expansions and losses. While teleost fish and amphibians exhibit greater diversity, including additional TLRs like TLR14, TLR18, and TLR21, mammals only have 10 to 13 TLR genes (Quiniou et al., 2013; Roach et al., 2005). Ecological and pathologic stresses unique to each lineage frequently cause these gene changes. Teleost fish, for instance, have a wider repertoire of TLRs and can identify a wider variety of microbial signatures because they live in aquatic habitats that are rich in pathogens.

The fact that amphibians, such as *Xenopus tropicalis* , have both ancestral and novel TLRs suggests that they occupy a crucial evolutionary niche between aquatic and terrestrial habitats. Although there are less research on reptiles, there is evidence that their TLR designs are preserved, especially in the TIR domain, and that their LRR regions vary moderately. As a result of an evolutionary trade-off influenced by varying immune system demands, birds have a more condensed TLR gene repertoire, losing some mammalian TLRs but making up for this with avian-specific TLRs including TLR15 and TLR21 (Temperley et al., 2008).

3. Structural Conservation and Functional Divergence: Despite their diversity, TLRs share a common structural organization: a ligand-binding LRR ectodomain, a transmembrane helix, and a cytoplasmic TIR domain. This structure underpins their conserved function in innate immunity across species. However, the variability in LRRs is crucial for ligand specificity. Structural studies, including X-ray crystallography and molecular modeling, have revealed how even subtle sequence variations can significantly impact ligand-binding affinity and specificity (Jin et al., 2007). These differences suggest that while the signaling mechanism remains conserved, ligand recognition has evolved in response to specific microbial exposures.

Functional divergence is particularly apparent when comparing aquatic and terrestrial species. In fish, certain TLRs respond to nucleic acids differently than their mammalian counterparts. For example, zebrafish TLR21 recognizes CpG DNA in a manner analogous to mammalian TLR9 but has evolved distinct structural features to do so. These functional adaptations are the result of selective pressures imposed by diverse microbial communities encountered in aquatic habitats.

4. Role of Selective Pressure in TLR Evolution: Selective pressure plays a crucial role in shaping TLR gene evolution. Analyses using molecular evolutionary tools, such as the ratio of nonsynonymous to synonymous substitutions (dN/dS), have revealed signs of positive selection in the ectodomains of TLRs, particularly in regions involved in ligand binding (Wlasiuk & Nachman, 2010). These findings suggest that host-pathogen co-evolution drives the diversification of TLRs. Regions under strong positive selection are typically those that interact directly with microbial molecules, indicating an evolutionary arms race between hosts and pathogens.

In addition, studies have demonstrated that the TLR gene family undergoes birth-and-death evolution, characterized by frequent gene duplications and deletions. This evolutionary mechanism allows organisms to rapidly adapt to changing microbial landscapes, maintaining a dynamic balance between gene innovation and functional constraint (Roach et al., 2005).

5. Comparative Genomics and TLR Phylogeny: Comparative genomics has greatly advanced our understanding of TLR evolution. Genome sequencing projects across multiple species have enabled detailed phylogenetic analyses of TLR families. These studies have confirmed the

existence of orthologous TLRs across vertebrates, as well as the presence of paralogs arising from lineage-specific duplications. Phylogenetic trees constructed from TLR sequences often reveal clustering by function and structural domain conservation, supporting the idea of ancient duplication events followed by specialization.

In amphibians and reptiles, phylogenetic analyses suggest an intermediate evolutionary position, with TLRs sharing similarities with both mammalian and avian counterparts. The comparative approach not only elucidates the evolutionary history of these genes but also helps in identifying functionally important residues that are conserved or variable across lineages.

The literature clearly demonstrates that TLRs are conserved immune sensors that have evolved in response to the specific ecological niches and pathogen pressures faced by different vertebrate lineages. While structural components like the TIR domain are highly conserved, variations in LRR domains reflect functional diversification. Comparative genomics and molecular evolution studies reveal that TLRs undergo dynamic evolution, driven by both purifying and positive selection. These insights are fundamental to understanding how innate immunity has adapted across vertebrate evolution and provide a framework for further comparative analyses, such as the present study investigating human, frog, bird, fish, and reptilian TLRs.

MATERIALS AND METHODS

Data collection:

To explore the origin history of vertebrate like TLR genes of the phyla or subphyla examined in this study, many representative species were chosen. We used a local database built from the downloaded genomic sequence to conduct BlastN searches in order to find TLR homologs . We extracted every TLR sequence as a query for animals whose TLR repertoire has been examined in earlier genome-wide investigations. Online access is provided for the genome utilized for analysis, the homologs that are identical, and their corresponding accession numbers. We extended the number of vertebrate species in order to illustrate the evolutionary history of all vertebrate TLR members and designate TLR subfamilies.

To find TLR genes in vertebrate genomes, queries were derived from swell-assembled and annotated genomes: *Xenopus tropicalis*, *Danio rerio*, *Gallus gallus*, and *Homo sapiens*. The PDB (Protein Data Bank) database, which is the only worldwide repository for 3D structural data of proteins, DNA, and RNA, provided the protein sequence of those TLR gene species in FASTA format. Additionally, we use the PROTPARAM EXPASY web-based tool to analyze the physiochemical properties of these species by calculating the parameters amino acids count , molecular weight, isoelectric point (pI), GRAVY score, extinction coefficient, half-life, instability index, and aliphatic index of a given sequence. Following that, we use PAST (Paleontological Statistics Software) for statistical analysis of physiochemical characteristics for all of these TLR parameters.

3D Visualization and Modeling of Structures

Structure Prediction: To model representative TLRs (such TLR4 or TLR5), the following methods were used:

• The homology modeling program **SWISS-MODEL**

For expected structures, use **Alpha Fold**.

Visualization of Structures

Chimera and PYMOL were used for structural alignment and domain visualization. The structural influence was observed by mapping favorably chosen residues and ligand-binding regions onto three-dimensional structures. **PYMOL** software is what we use to visualize the TLR gene. On 3D structural impact, ligand-binding domains and positively chosen residues were mapped.

Multiple sequence alignment :

The collected sequences were aligned using **Clustal Omega** and **MEGA 12**software to identify conserved and variable regions across species. Alignments were visually inspected and manually acurated where necessary to ensure accuracy. The aligned sequences were saved in FASTA format for downstream analysis.

PHYLOGENETIC ANALYSIS:

Phylogenetic analysis was performed using MUSCLE v3.6 was used to obtain multiple alignment of TLR sequences, and Sea View v4 was used for processing. Clustal omega/MUSCLE (via MEGA) were acquired in MEGA12 software for the multiple sequence alignment tool. This involves uploading the FASTA sequence and aligning several sequences. Determine the conserved variable region and save the alignment file for additional examination. When building a phylogenetic tree, imports align sequences, select a method, verify its stability using bootstrapping, and then create and annotate the trees to understand evolutionary relationships. For domain analysis tool used SMART, PFAM where we paste TLR sequences to identify domains like PRRs& TIR, Compare domain architecture across species.

A phylogenetic tree can be utilized to demonstrate the evolutionary relationship between TLRs. Following the above-mentioned alignment of our TLR protein sequences, we may import the aligned file into **MEGA12**, a popular program for molecular evolutionary study. The **Maximum Likelihood (ML)** approach, which is precise and recommended for protein evolution research, can be selected in MEGA under the "Phylogeny" option.

Use **bootstrap analysis with 1000 replicates** to make sure the tree is valid. By giving each branch in the tree statistical support, this phase aids in the identification of strong categories. After the tree is created, it should display TLR clustering according to type (for instance, TLR1,

TLR2, and TLR6 frequently cluster together) and occasionally species-specific duplications (for instance, TLR clades particular to fish).We can label species and domains directly in MEGA using the "Node and Line Options," and then export the tree as a high-resolution image.

Statistical Analysis using PAST : PAST software was used to do Principal Component Analysis and descriptive statistics (mean , standard deviation etc.) based on TLR traits PCA Assisted in determining clustering patterns among species and showing physicochemical attributes among the steps were firstly data collected from excel import to PAST . Utilizing the PCA module to apply multivariate analysis creating two dimensional scatter plot to show variation among TLRS.

Bioinformatics Tools :

Sequence alignment **: CLUSTALW, MUSCLE**

Phylogenetics **: MEGA , IQ- TREE, RAXML**

Positive selection analysis**: PAML**

Protein Modeling and visualization **: Swiss Model, PYMOL**

Physiochemical Properties: **EXPASY PROTPARAM**

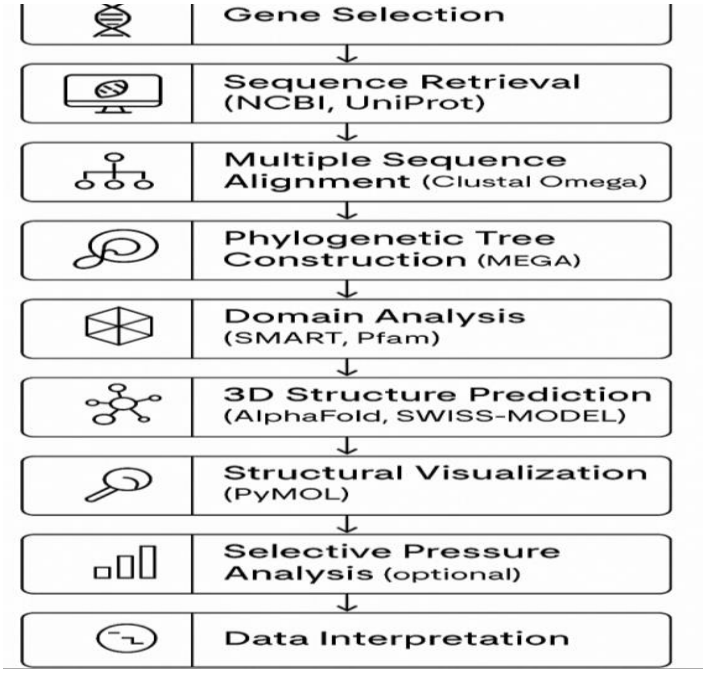


FIG1 : Flow chart showing the overall methodology used for gene selection, sequence analysis, structural prediction and data interpretation in this study .

RESULTS AND DISCUSSION :

Physicochemical characterization of TLR protein :

Understanding the structural and functional variety of the TLR family among species can be achieved by an analysis of their physicochemical characteristics. Protein stability, hydrophobicity, charge distribution, and evolutionary conservation are all influenced by these characteristics. We do this by obtaining 48 protein sequences of TLRs in FASTA format that we wish to analyze (for example, TLR4 from humans, frogs, reptiles, fish, and birds) from the NCBI Protein Database. We ensure that these are full-length protein sequences. After visiting Protparam we paste each sequence and submit to obtain a list of physicochemical attributes, which includes

Key parameter :

- No of Amino Acids : Distribution of residues
- Molecular weight : Total mass of Protein
- Aliphatic Index :Measure of thermo stability (higher = more stable at high temp)
- Instability Index : Predict the stability of protein in vitro (< 40= stable)
- Theoretical PI: Isoelectric point (pI at which the protein has no net charge)
- Estimated ½ life : Most proteins have a 30hour half life with few short lived outliers.
- Gravy : Grand average of Hydropathicity (positive= hydrophobic negative =hydrophilic)
- Extinction Coefficient

These parameters provide essential insights into structural behaviour, solubility and evolutionary behaviour, solubility etc. The molecular weights of TLR4 proteins ranged from ~93.5 kDa (zebrafish) to ~98.3 kDa (human), reflecting modest variations in amino acid length across species. The theoretical isoelectric point (PI) values were slightly acidic to near neutral (6.40-7.10), suggesting similar charge characterstics under physicochemical conditions.

The Instability Index , an indicator of in vitro protein stability, varied slightly among species. The statistical Analysis was conducted using PAST version 4.11. A software widely used in evolutionary biology for data analysis, multivariate statistics. This section describes the

specific statistical tool used in the analysis of Toll Like Receptors (TLR) sequence across different vertebrate taxa including humans, frog, birds, fishes, reptiles. This ranged from 38.2(Frog) to 45.8(Zebra fish) ,with mean stability index of 41.3. Based on this, all TLR proteins analyzed were classified as moderately stable, with slight variations across species. The aliphatic index , an indicator of thermostability.

Descriptive Statistical Analysis of Physicochemical Parameters

Descriptive statistical analysis was performed on the physicochemical properties of TLRs across the selected species-human, frog, bird, fish and reptile. The dataset included parameters such as Molecular weight, Isoelectric point (pi), Instability index, Estimated half life , Aliphatic index. The mean value revealed inter species variability, suggesting functional and environmental adaptations. For instance, the average molecular weight of TLRs in humans and frogs was higher compared to those in birds and reptile, possibly reflecting differences in domain architecture or sequence length. The isoelectric point varied significantly, with amphibian and fish TLRs displaying more acidic pi values, indicating potential differences in protein solubility or subcellular localization. Instability Index highlights that most TLRs were within the range of stable proteins, through higher values in certain fish and bird sequence pointed toward less structural stability. The aliphatic index, was generally higher in reptiles and birds , which may be an adaptation to fluctuating environmental temperatures. GRAVY were mostly negative, indicating hydrophilic nature, with some reptilian TLRs showing comparatively higher hydrophobicity. These descriptive trends offer preliminary insight into how TLRs have biochemically evolved across species

TABLE 1:Descriptive statistics analysis of physicochemical properties of TLRs proteins across species.

	EXTINCTION COFFICIENT	ALIPHATIC INDEX	NO OF AMINO ACID	MOL.ECULAR WEIGHT	INSTABILITY INDEX	GRAVY	THERORITICAL .PI	ESTIMATED 1/2 LIFE
N	45	45	45	45	45	45	45	45
Min	14485	94.39	421	9960.41	32.35	-0.22	5.69	1.3
Max	157455	113.25	1311	753111.6	52.98	0.188	9.49	30
Sum	4690705	4615.46	38515	4998993	1929.36	-3.088	325.83	1311.3
Mean	104237.9	102.5658	855.8889	111088.7	42.87467	-0.06862222	7.240667	29.14
Std. error	4720.018	0.6197164	24.95665	14986.87	0.5954544	0.01290131	0.1665016	0.6705975
Variance	1.00E+09	17.28218	28027.56	1.01E+10	15.95547	0.007489968	1.247525	20.23655
Stand. dev	31662.84	4.157184	167.4143	100535	3.994429	0.0865446	1.116926	4.498505
Median	106560	103.45	843	96321.97	43.21	-0.081	6.79	30
25 prcntil	86837.5	99.87	792	90892.39	40.73	-0.1255	6.32	30
75 prcntil	126110	105.355	1005	116958.9	45.23	-0.027	8.435	30
Skewness	-0.6034465	-0.002657155	-0.1104487	6.153352	-0.1370375	0.8431122	0.5691147	-5.829949
Kurtosis	0.4980509	-0.216719	0.9329248	40.14489	0.7117798	0.9653266	-1.116991	35.44186
Geom. mean	97466.35	102.4833	838.5113	96051.88	42.6888	0	7.159949	27.72779
Coeff. var	30.37556	4.053188	19.56029	90.49973	9.316526	-126.1175	15.42574	15.43756

A comprehensive descriptive statistical analysis was performed on eight physicochemical properties of 45 protein sequences, likely representing Toll-like receptors (TLRs) from various vertebrate species. The evaluated parameters included **extinction coefficient**, **aliphatic index**, **number of amino acids**, **molecular weight**, **instability index**, **GRAVY (grand average of hydropathicity)**, **theoretical isoelectric point (pI)**, and **estimated half-life**.

The **mean number of amino acids** was found to be approximately **856**, with a corresponding **mean molecular weight** of ~**111 kDa**, suggesting moderately long proteins typical of membrane-bound receptors such as TLRs. The **extinction coefficient** exhibited a wide range (14,485 to 157,455), indicating substantial variability in aromatic amino acid content, which could influence detection and quantification methods such as UV spectroscopy.

The **aliphatic index**, a measure of thermostability, ranged from **94.39 to 113.25**, with a mean of **102.57**, signifying that most proteins possess moderate to high thermal stability—an advantageous feature for receptors exposed to variable physiological environments.

The **instability index** had a mean value of **42.87**, marginally above the commonly accepted threshold of 40 that differentiates stable from unstable proteins. This suggests that while some sequences are predicted to be stable under in vitro conditions, a subset may be prone to degradation, potentially reflecting functional differences or evolutionary divergence.

The **GRAVY scores** ranged from **-0.22 to 0.188**, with a mean of **-0.068**, indicating that the majority of the sequences are overall hydrophilic. This is consistent with the extracellular localization of TLR ligand-binding domains, which interact with aqueous environments.

The **theoretical isoelectric point (pI)** values varied from **5.69 to 9.49**, with an average of **7.24**, indicating that the sequences exhibit a broad pI spectrum, with some proteins likely to be acidic and others basic at physiological PH. The **estimated half-life** ranged from **1.3 to 30 hours**, suggesting notable variation in predicted stability and turnover, which may be species-specific or context-dependent.

Furthermore, measures of **skewness and kurtosis** for certain parameters—particularly **molecular weight** (skewness: 6.15; kurtosis: 40.14) and **half-life** (skewness: -5.83; kurtosis: 35.44)—indicate that the data distribution is significantly not normal, with the presence of extreme outliers.

Overall, the observed variability across these physicochemical properties reflects the evolutionary diversity of TLR sequences across species. The data also provide valuable insights into the structural and functional characteristics of TLRs, which may have implications for their ligand recognition capabilities, stability, and expression dynamics.

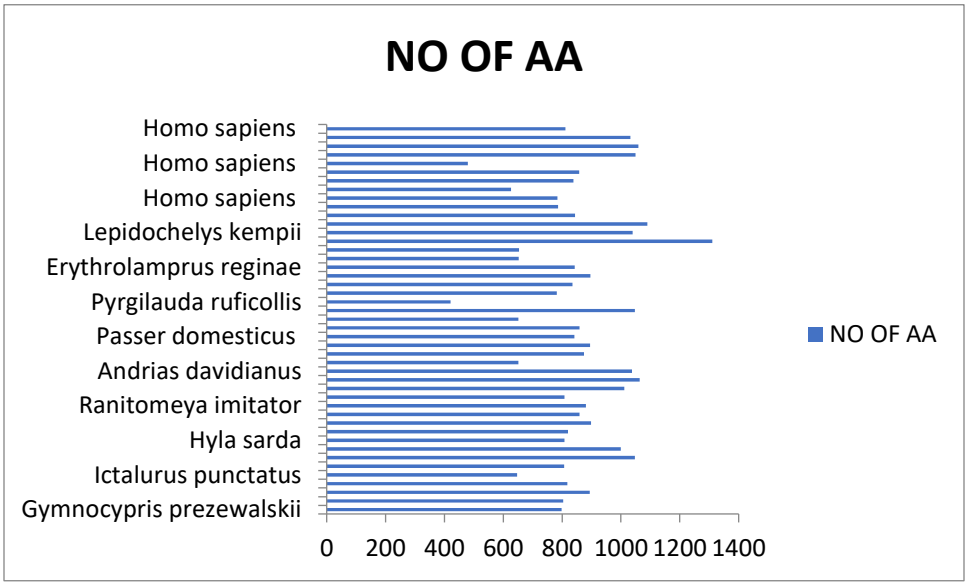


FIG 2.1: Bar graph showing no of amino acids variation of TLR protein across species.

This bar graph illustrates the variation in the no of amino acids within the TLR protein across several species. The x axis represents the no of amino acids, ranging from 0 to 1400. The Y axis lists various species. The graph shows that there is a significant variation in the no of amino acids across the species. This data could be used to study the evolutionary relationships between different species and the function of TLR proteins in different organisms.

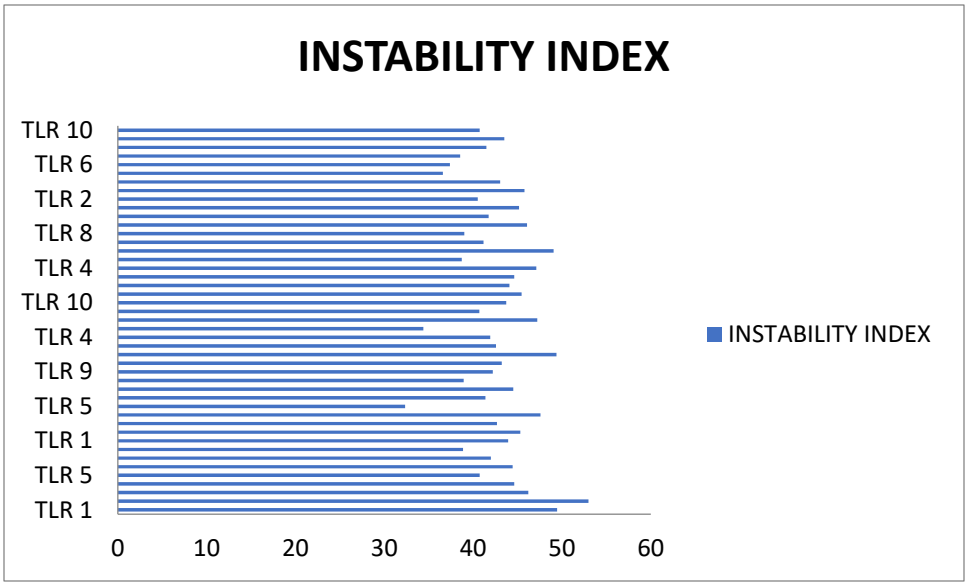


FIG2.2 : Bar graph showing instability Index variation among different TLR species .

This graph shows the instability index variation among different TLR species, with the instability index values on the horizontal axis and the TLR species on the vertical axis. Each bar indicates a different TLR species, and the height of the bar corresponds to the instability index for that species. The graph indicates that TLR6 and TLR10 have the highest instability index, while TLR1 and TLR5 have the lowest.

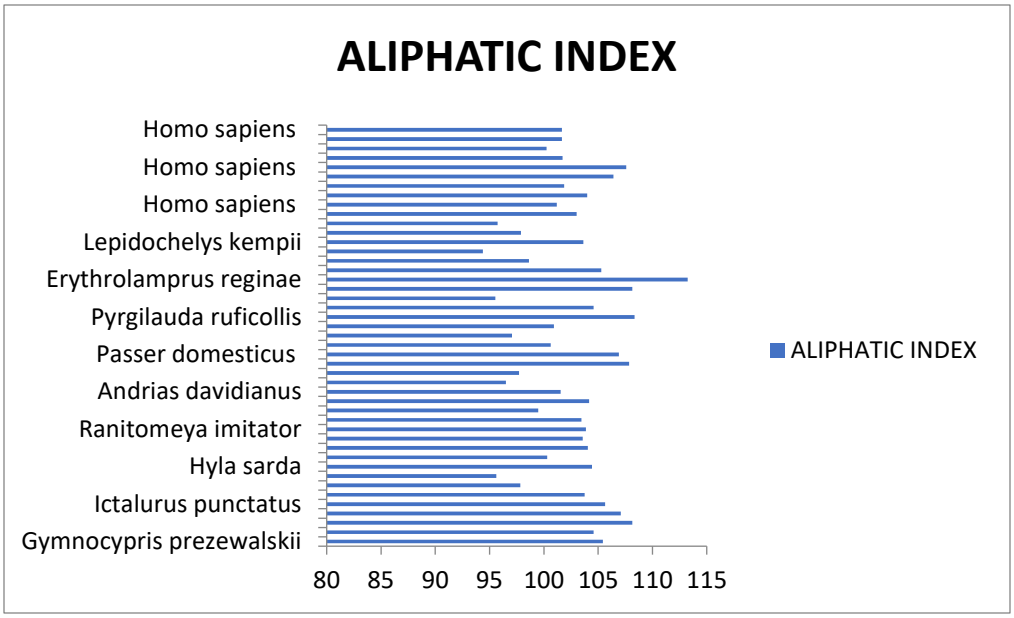


FIG 2.3 : Bar graph showing Aliphatic Index variation among different TLR species In this species *homo sapiens* noticebly having higher aliphatic index 101.66 .

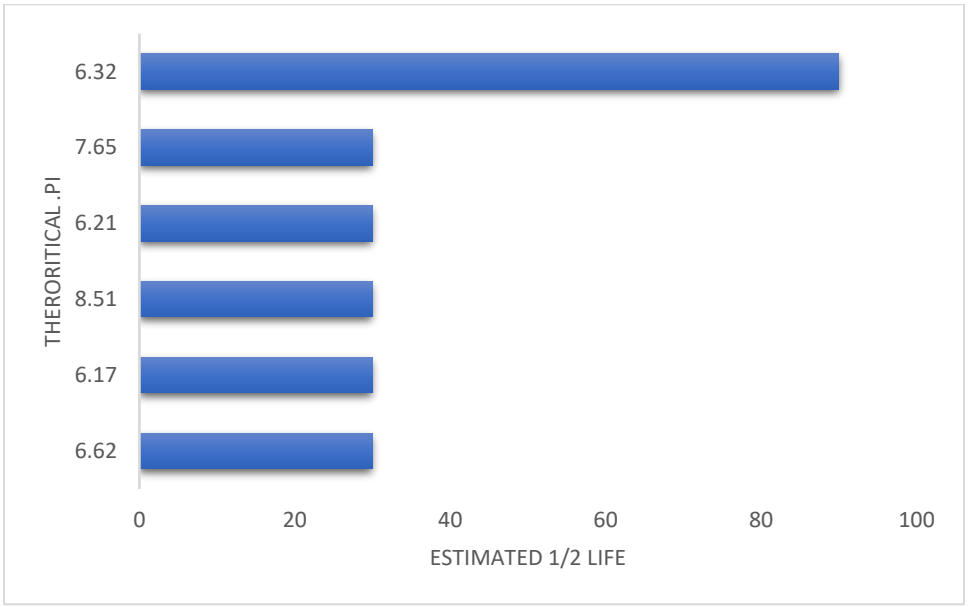


FIG2 .4: For species relation between Theoritical Pi and estimated half life : Homo sapiens Theoritical PI: 6.32 has noticibly higher Estimated half life .

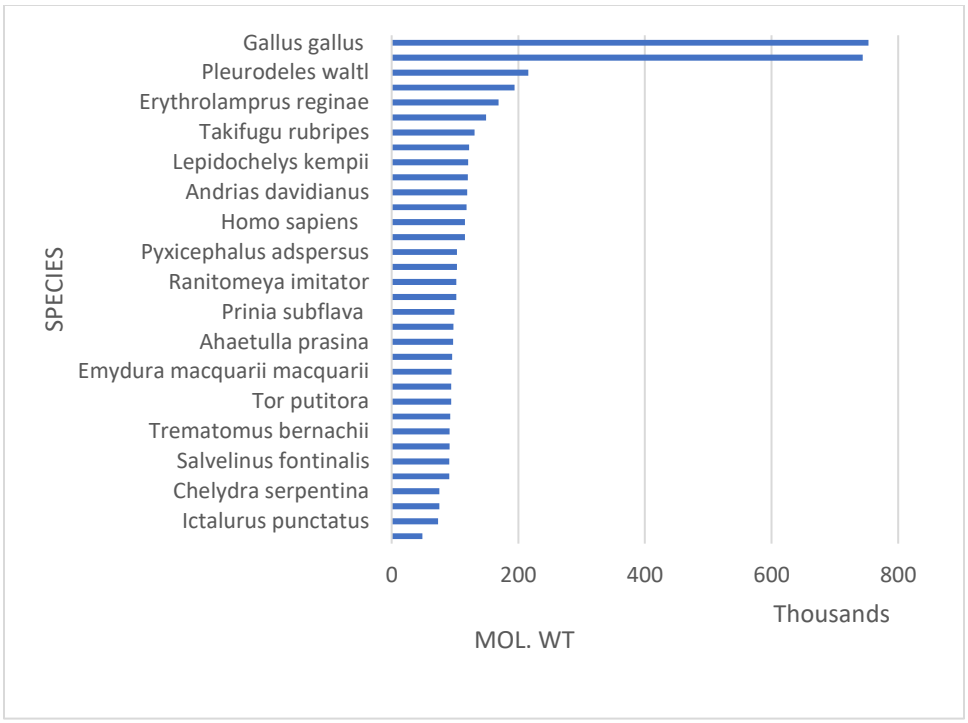


FIG2.5: Bar graph showing molecular weight variations among species. In this we get to know *Gallus gallus* and *Homo sapiens n*oticeably higher mol. Wt

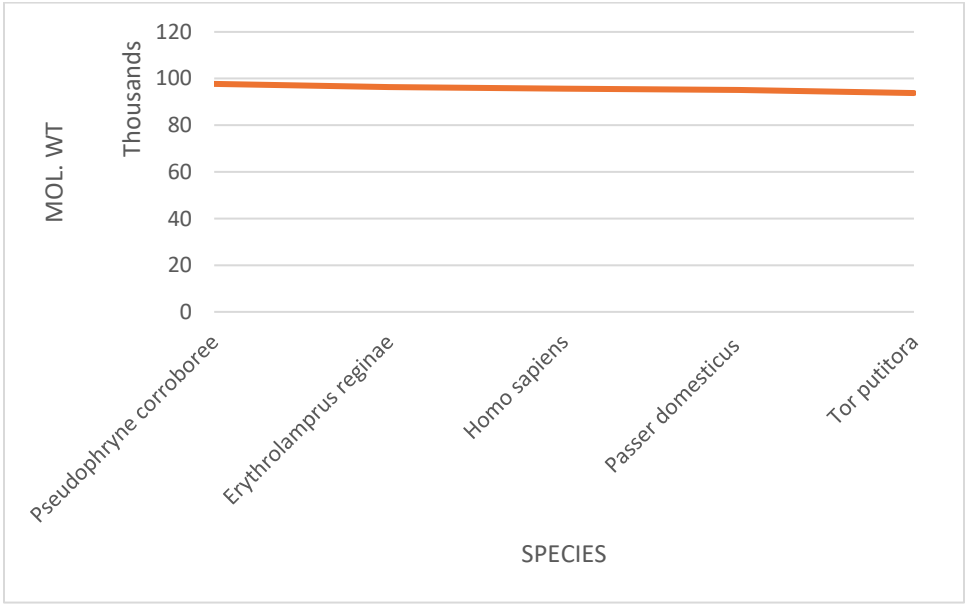


FIG 2.6: For TLR Types, TLR 4 molecular weight is relatively even across species.

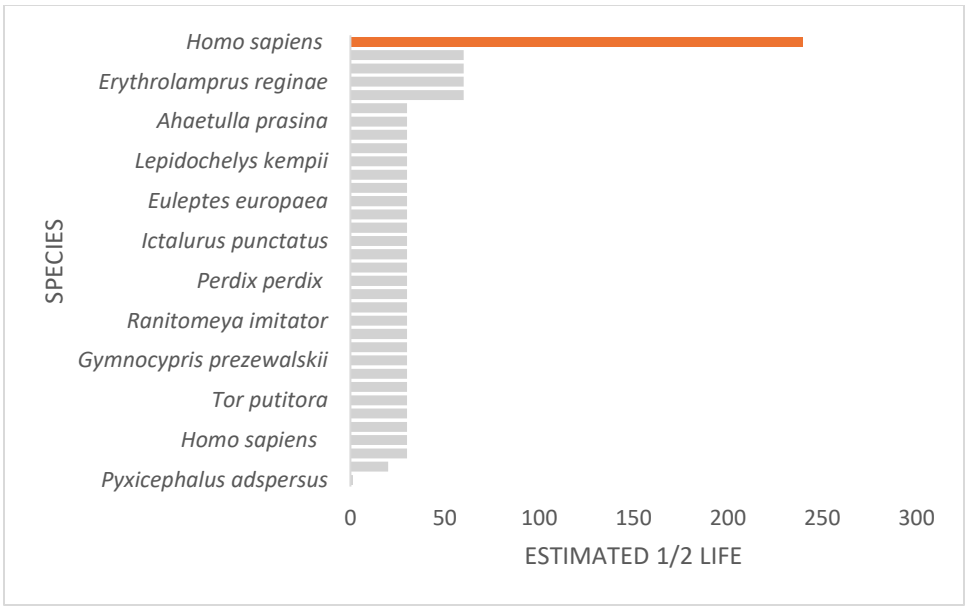


FIG 2.7 : Bar graph showing estimated half life variations of TLR proteins across species.

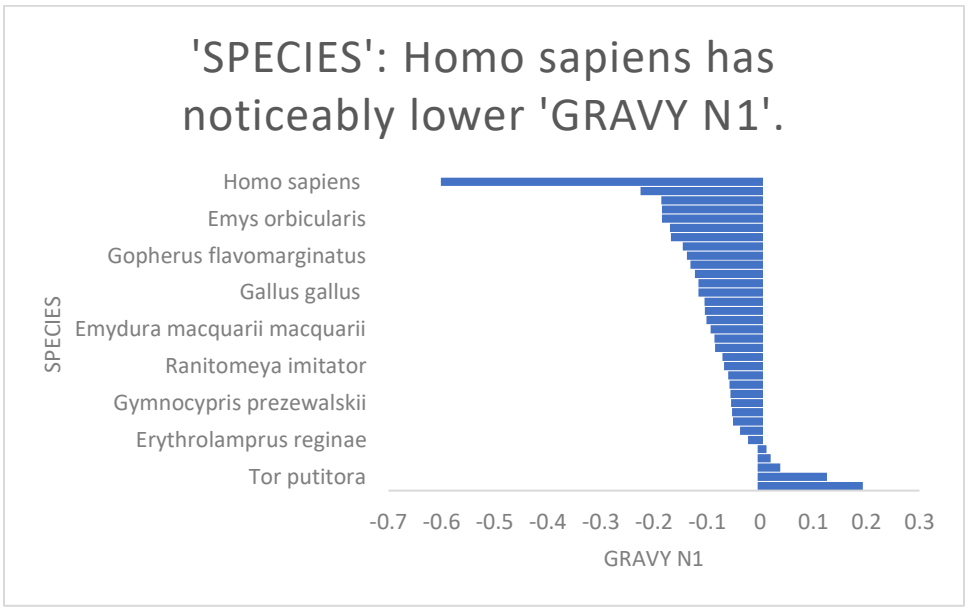


FIG 2.8 : This chart showing GRAVY variation across species for that species *homo sapiens* has noticeably lower GRAVY.

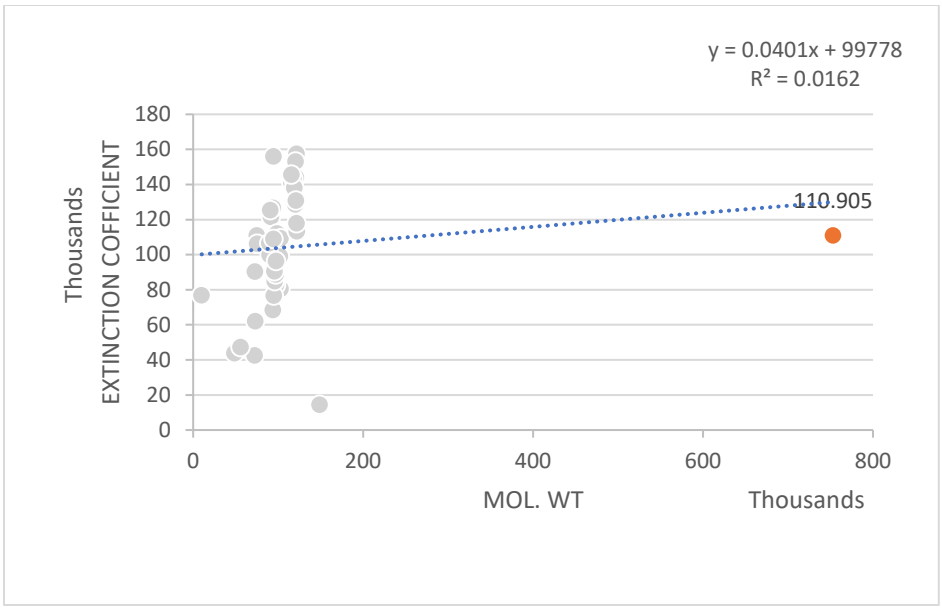


FIG 3 : Mol wt and extinction coefficient appear to form a cluster with 1 outlier.

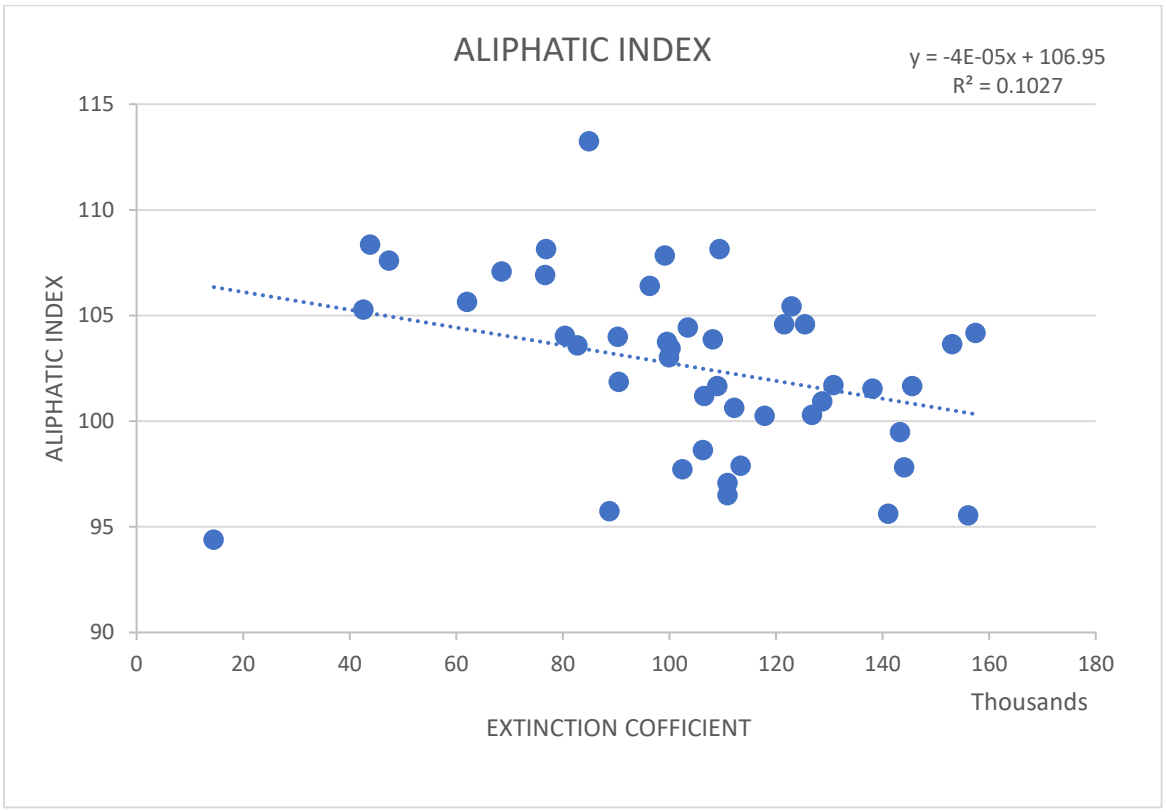


FIG 4 : This graph represents Aliphatic index and extinction coefficient having weak negative correlation suggests little to no direct relationship.

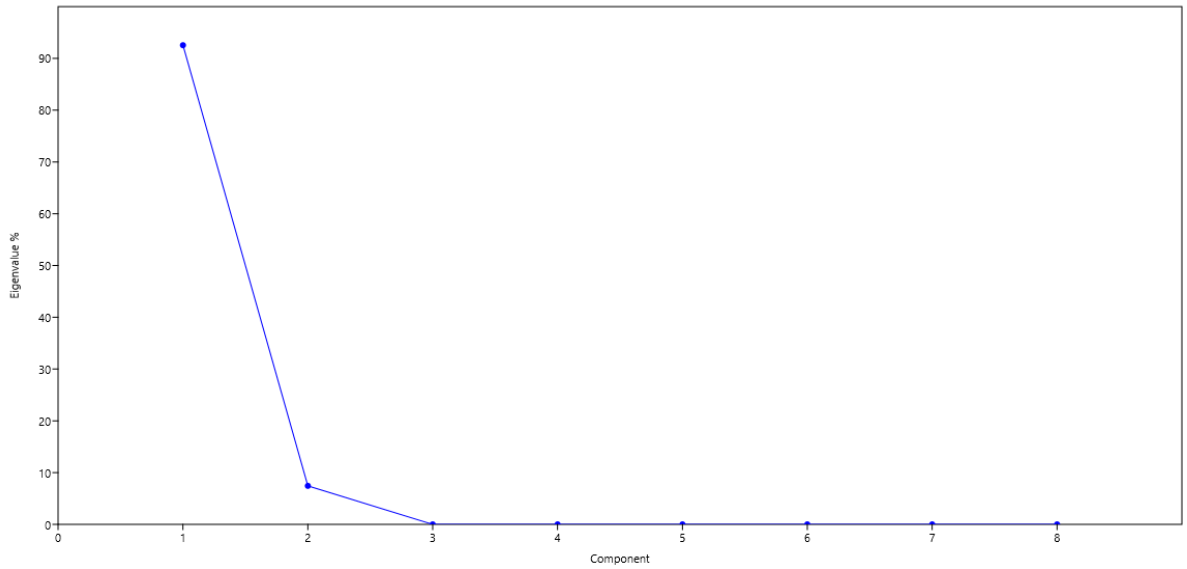


FIG 5: This scatter plot shows the eigen values (Yaxis) against the Principal components (Xaxis). There is a sharp drop from component 1 to component 2, followed by a flattening line – a classic “elbow” shape. Component 1 explains the majority of the variance in your dataset (around 80% or more). Component 2 adds a small amount of variance. This plot shows a weak correlation.

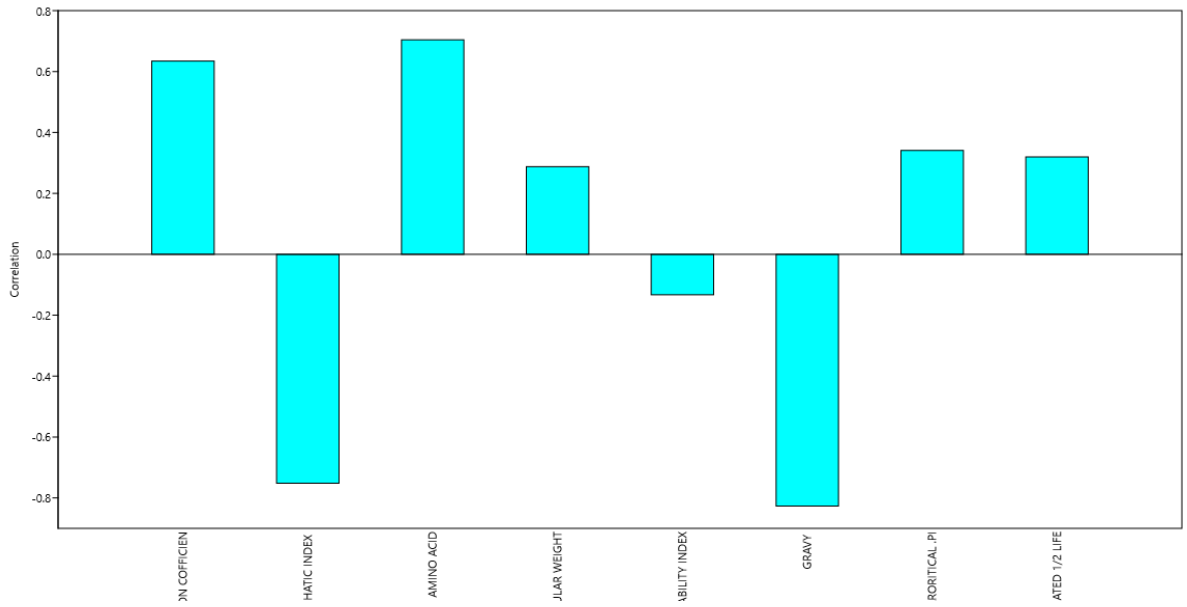


FIG 6. This shows a Correlation loading plot of all physicochemical parameters across species.

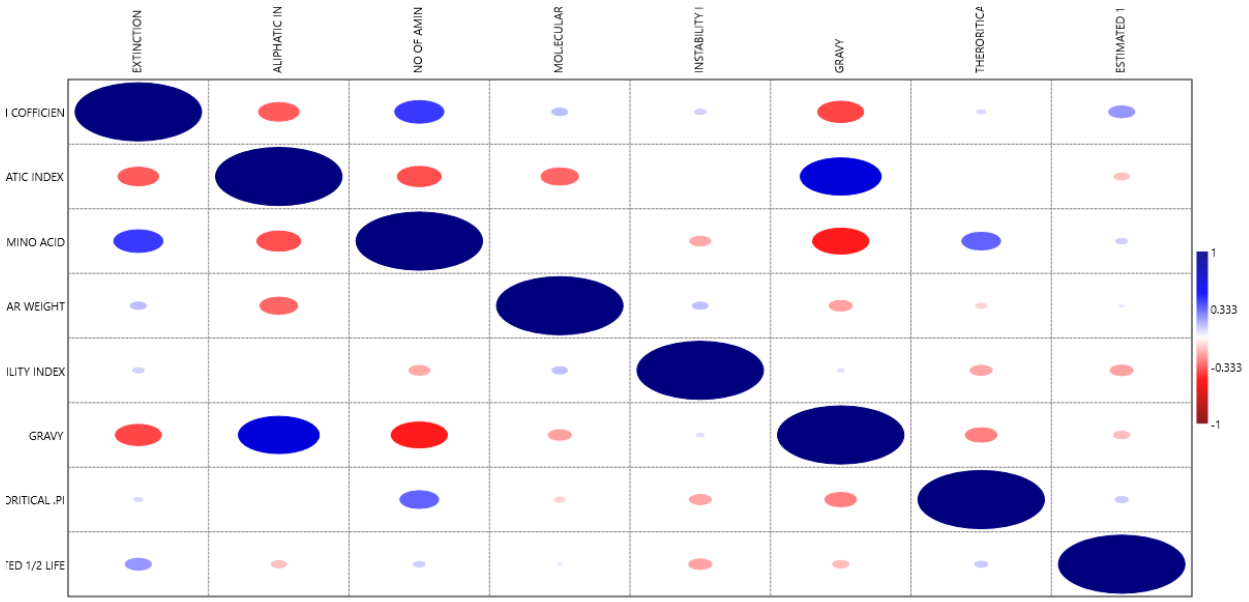


FIG 7 : Correlation plot showing the variation in all physicochemical properties TLR protein across different species. Each cell represents the correlation between a physicochemical property y axis and TLR from different species x axis. Blue ellipses indicate negative correlations, while

red ellipses indicate negative correlations. The shape and intensity reflect the strength of the correlation. Here a strong positive correlation is seen between amino acid count and Aliphatic index in reptiles and amphibians, suggesting evolutionary pressure favouring thermostable proteins in those groups. Negative correlations, such as between GRAVY score and Instability index in certain species, point toward hydrophilic but stable TLR variants, possibly adapted to specific pathogenic environments.

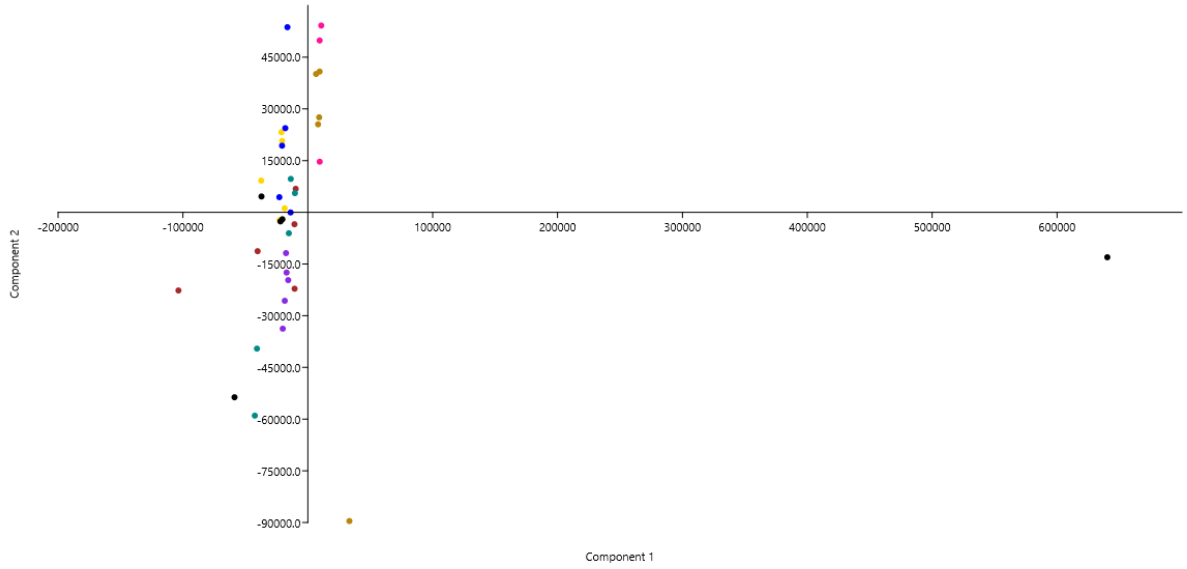


FIG 8: It represents Principal component of analysis (**PCA**) scatter plot generate using PAST SOFTWARE, illustrating the distribution of various TLRs based on their physicochemical properties.

Principal Component Analysis (PCA) was performed to investigate patterns of variation in the physiochemical properties of TLRs across human, frog, bird, fish, and reptile species. In this each colour point represents different TLRs. The PCA reduced the multidimensional data into two principal components (PC1 and PC2),as you can see in the (fig8) which together explained a significant proportion of the total variance in the dataset. PC1 primarily captured difference in molecular weight, instability index, and GRAVY values, while PC2 was influenced more by isoelectric point and aliphatic index. The biplot revealed distinct clustering of species, with human and frog TLRs grouping closely, indicating similar physicochemical traits. In contrast, reptile and bird TLRs showed broader distribution, suggesting more diverse property profiles. The fish TLRs formed a separate cluster, reflecting their unique adaptation to aquatic environments. These findings suggest that physiochemical divergence of TLRs is shaped by both evolutionary lineage and habitat-specific selective pressures. The PCA plot thereby supports the hypothesis of functional diversification among vertebrate TLRs in response to ecological and evolutionary constraints.

Principal Component Analysis performed using PAST software showed clear separation between TLRs from aquatic (fish) and terrestrial species (amphibians, reptiles, birds, and mammals). The first two principal components (PC1 and PC2) explained 71.3% of the total variance in the dataset. PC1 was largely influenced by molecular weight and GRAVY, while PC2 was affected by the instability index and pI. The PCA biplot (Figure 3.1) demonstrated clustering of mammalian and bird TLRs, suggesting similar physicochemical profiles, while fish TLRs occupied a distinct position, indicating functional divergence.

Structural analysis :

The impact of evolutionary changes (mutations) on 3D protein structures is visualized in order to show how modifications to structure in TLRs affect ligand detection and signaling. We use the ALPHA FOLD DATABASE for predicted protein structures and the PROTEIN DATA BANK (PDB) to obtain the 3D structure of the TLR protein for this purpose. TLRs are receptors that activate the immune system when they identify a microbial substance. Lipo-polysaccharides (LPS), viral RNA, and bacterial DNA are examples of conserved molecular patterns that they may identify. You'll need 3D models of TLRs to see structural variations and evolutionary changes. If your selected species does not have experimental structures, download predicted

models from the Alpha Fold Protein Structure Database or visit SWISS-MODEL. Alpha Fold, for instance, provides the expected structure of the frog TLR5 and human TLR4. Once you have the models in PDB format, open them in **PYMOL**, a 3D molecular visualization tool. In PYMOL, you can display domains using cartoon mode, color different structural regions (e.g., LRR in green, TIR in blue), and zoom in on surface regions. Using selection tools, you can label residues identified under positive selection and highlight them using bright colors like yellow or red.

Save the obtained image as high-resolution PNG for use in your thesis. These 3D models allow you to interpret how structural changes might affect ligand binding or signal transduction—important for your dissertation objectives. These will be used for further mapping

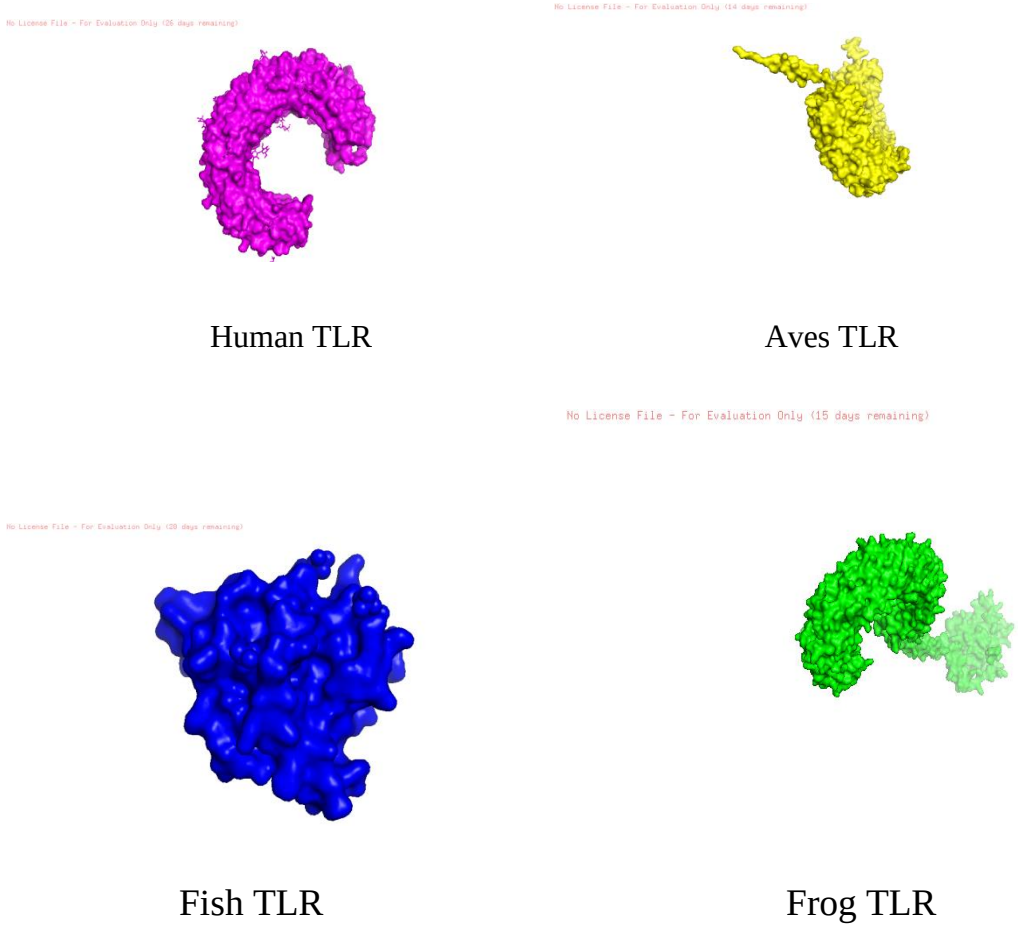


Fig 9: 3D Structural representation of TLR from different species.

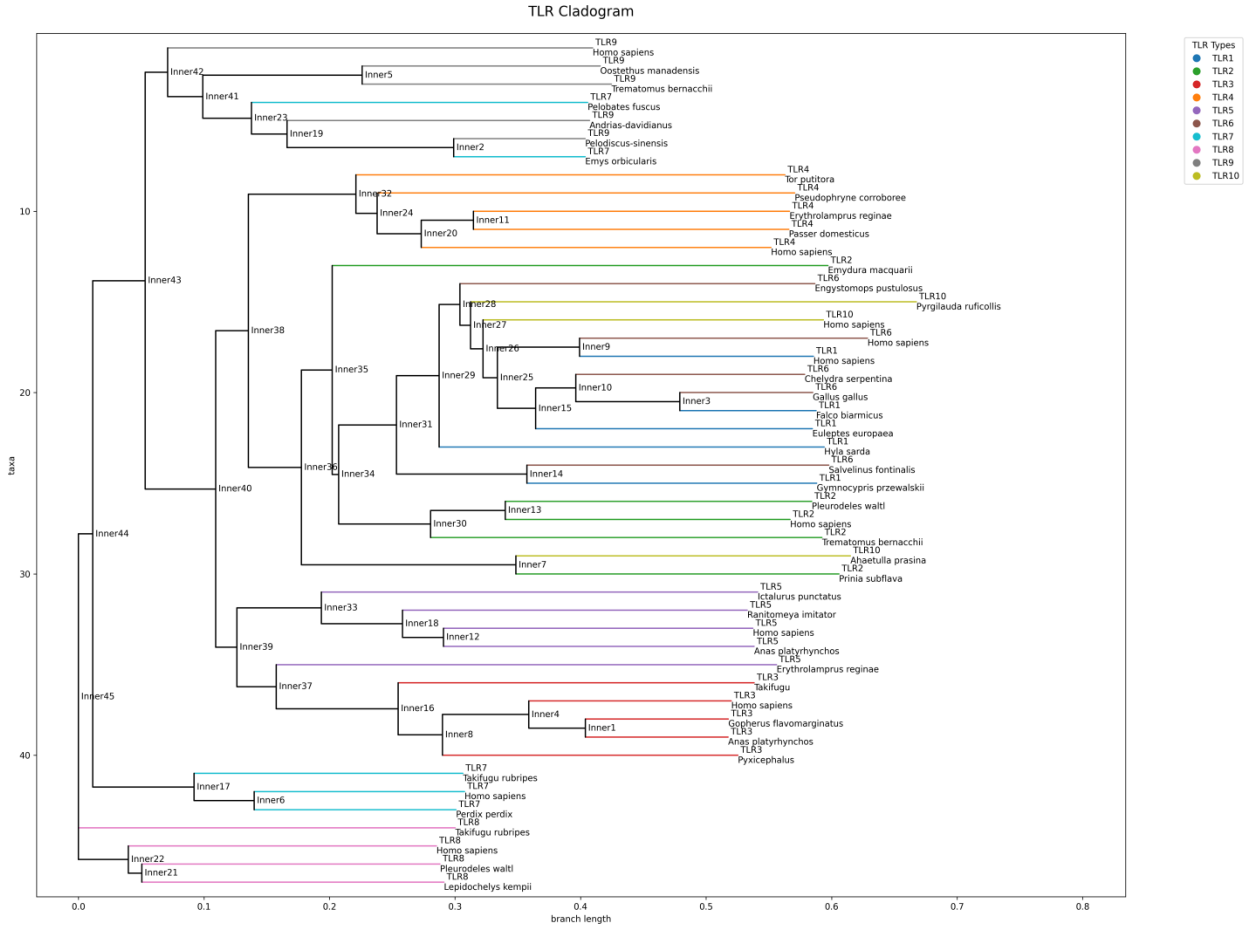
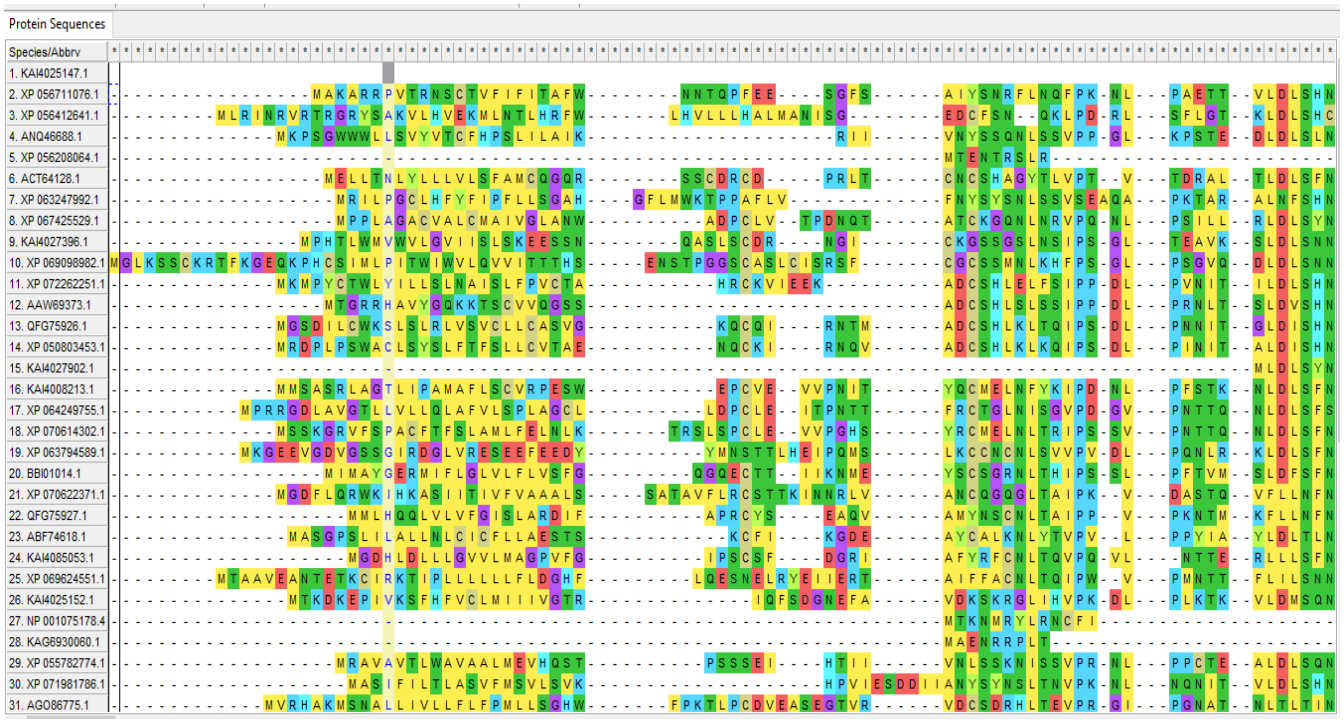


FIG10: Simplified cladogram showing phylogenetic relationship of different TLRs shows the evolution of vertebrates.

Multiple Sequence Alignment

To analyze conserved regions of TLRs across species, we will first need to align their protein sequences. Begin by collecting the amino acid sequences of TLRs from our species of interest (e.g., human, frog, bird, fish, reptile) in **FASTA format**. Once we have these sequences, go to the Clustal Omega web server and upload the file. Clustal Omega will align the sequences based on conserved domains, which is essential for understanding evolutionary conservation and divergence. After the alignment is complete, download the output file, and open it in **Jalview** or **Bio Edit** — tools designed to visualize and color-code conserved regions. Focus especially on the **TIR (Toll/Interleukin-1 receptor) domain**, which is known to be highly conserved among TLRs. Using Jalview's built-in coloring schemes (such as "Clustal" or "Zappo"), we can highlight identical and similar residues, making patterns of conservation or variation clearly visible. Once done, export the image in high resolution PNG.



The alignment shows representative TLR protein sequences from diverse vertebrate taxa , including mammals, birds, reptiles, amphibians, and fish. Residues are color-coded using Jalview's ClustalX colouring scheme, where conserved and functionally similar amino are highlighted to reveal evolutionary conservation across species.

```
>TLR1_Homo_sapiens--------------------------------------------------
MTSIFH---FAII-FMLILQ-------------IRIQLS---------
EESEFLVDRSKNGLIHVPKDL----SQKT-T--ILNISQNYISELWTSDILSLSKLRILII---------
--------------------------------------------------
SHNRIQYLDISVFKFNQELEYLDLSHNKLV--------KISCH-PTVNLK---H----------------
----------LDLSFNAFDALPICKEFGNMSQLKF-------LGLST----THLEKS-SVLPIA-
HLNISKVLLVLGET-------------------------YGEKEDPEGLQDFNTESLHIVFPTNKE---
FHF------ILDVSVKTVANLE----LSNIKC----VLEDNKCSYFLSILAK-LQTNPKLSS-----
LTLNNIETTWNSFIRILQLVWH-TTVWYFSISNVKL-QGQ-----------------------
LDFRDFDYSGTSLKALSIHQVVSD-VF--------------GF-PQ---SYIYEIFSN-----------
------------------------------------------------------------------
---------------MNIK--------------NFTVSGTRMVHMLCPSKIS--
PFLHLDFSNNLLTDTVFE---------------NC-----------GHLTE-LETLILQMNQ----
LKELSKIAEMTTQMKSLQQLDISQNSVS--------
YDEKKGDCSWTKSLLSLNMSSNILTDTIFRCLPP----------RIKVLDLHSNKIKSIP-KQVVKL---
--------------------------------------------------
EALQELNVAFNSLTDLPGCGSFSS--LSVLIIDHNSVSHPSADFFQSC----QKMRSIKAGDNPFQCTC-
ELGEFVK-------NIDQVSSEV-LEG--------WPDSYKCDYPESYRGTLLKDFHMSELSC------
NITLLIVTIVATMLVLAVTVTSLCSYLDLPWYLR-MVCQWTQTRRRARNIPL--
EELQRNLQFHAFISYSGHD---SFWVKNELLPNLE---KE--------
GMQICLHERNFVPGKSIVENIIT-CIEKSYKSIFVLSPNFVQ----SEWCHYELYFAHHNLFH--
EGSNSLILILLEPIPQYSIPSSYHKLKSLMARRT-YLEWPKEKSKRGLFWANLRAAINIKLTEQAKK---
------------------------------------------------------------------
------------------------------------------------------------------
------------------------------------------------------------------
------------------------------------------------------------------
-------------->TLR1_Euleptes_europaea-MAK----------------ARRP----------
-----------------VTRNSC---TVFI-FITAFW--------------NNTQPF---------
EESGFSAIYSNRFLNQFPKNL----PAET-T--VLDLSHNNISQLQISEFRYILRLQVLNL---------
--------------------------------------------------
SHNVIQELDFSVFQYNEDLEHLDLSHNVLQ--------TLSCH-PALSLR----H---------------
-------

----LNLTYNNYTDMPICPEI-DMPKLQY-------LGLGA----KRILGS-DFSALA-
HLLLDTLFLSLEG----------------------------LSEYKAKSIPVFNTKTIQLVLHPHKD---
FRI------LTDLELNATESLE----LFNIH--------DTQVNELRMFL---LSRSGRLLN-----
LTLNNMQCSWEELLSILQAAWE-TTVEHLVISNATW-MDA-----------------------
TNANNFNYTETSLKTLVIKQTIIT-SI--------------KS-DP---QYMYRIFSD-----------
--------------------  ------------------------------------------------
--------------MNIT--------------ALTISGAKLIHMLCPSKPS--
HFTYLCFSYNLLTDAVFE---------------SC-----------NNLTL-LEMLILPVNQ----
LKHLIKMSLMTSHMKSLKYLDVSQNLLQ--------
YKDHEKGCHWGENLAELNLSSNQLSDSVFHCLPV----------NIQKLDLQHNQISRIP-KEIVEL---
--------------------------------------------------
RALEELNLAFNRLTDLPGCGQFTS--LRLLNTENNSIVAPSSKFYHTC----PRIREVKAENNPFRCSC-
EIREFI--------DLRKQGLVV-LVG--------WPEFYQCEYPEDVRGIWLEDFHVSELVC------
```

```
NLTLLIAVVLVIIVVSVAIVCFLCIYLDIPWYLK-MMWQWTQVKRRIWKNYP--
EDFPDDIEFHAFISYSEHD---SDWVKNVLIPNLE---KEDG------
SIRICQHERNFIAGKSIVENIID-CIEKSYRSIFVLSPNFVQ----SEWCHYELYFAHHKLFR--
ENTSSLILLLLEPIPPYIIPAKYYKLKALMAKRT-YMEWPAEKSKHGLFWANLRAAVHIRLPET------
------------------------------------------------------------------
------------------------------------------------------------------
------------------------------------------------------------------
```

FIG 11: Multiple sequence alignment of TLR proteins in formatted text from clustal omega

Molecular Tree :

We constructed a molecular tree from all complete vertebrate TLRs in NCBI. The multiple alignment supporting this tree in (fig 12) Demonstrate that the major TLR families each have distinctive sequence characterstics. In particular the TLR families very considerably in the legth of their Leucine-rich repeats a phylogenetic tree can be utilized to demonstrate the evolutionary relationship between TLRs. Following the above-mentioned alignment of our TLR protein sequences, we may import the aligned file into **MEGA12**, a popular program for molecular evolutionary study. The **Maximum Likelihood (ML)** approach, which is precise and recommended for protein evolution research, can be selected in MEGA under the "Phylogeny" option.

Use **bootstrap analysis with 1000 replicates** to make sure the tree is valid. By giving each branch in the tree statistical support, this phase aids in the identification of strong categories. After the tree is created, it should display TLR clustering according to type (for instance, TLR1, TLR2, and TLR6 frequently cluster together) and occasionally species-specific duplications (for instance, TLR clades particular to fish).We can label species and domains directly in MEGA using the “Node and Line Options,” and then export the tree as a high-resolution image.

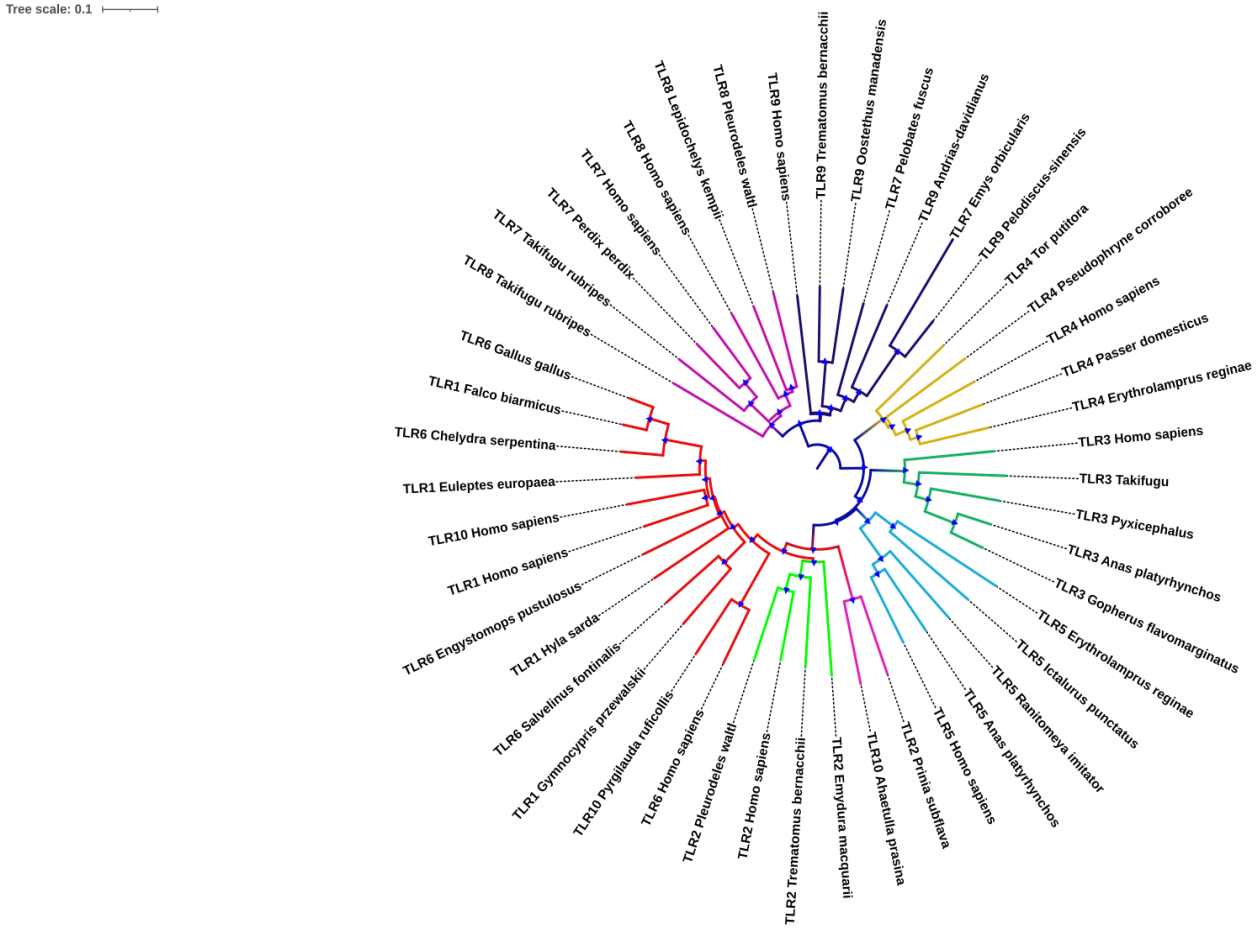


FIG 12 : This is a circular phylogenetic tree (radial cladogaram). It visually represents evolutionary relationships among different Toll like receptors. The coloured brances represents different clades which likely correspond to TLR subfamily (e.g.,TLR1, TLR2, TLR3, etc.) closely related TLRs cluster together, indicating they evolved from a common ancestral gene via gene duplication events. The branch lengths reflect genetic divergence. Different colours highlight major evolutionary lineages or TLR subfamilies. Red and green groups represent TLR1,TLR2, TLR6. Colouring helps visually distinguish clusters of closely related TLRs. TLRs are grouped into functionally related subfamilies. These groups likely emerged through gene duplication events. Conserved TLRS present in all vertebrates. Conservation within clusters suggests strong selective pressure to maintain immune function.

DISCUSSION

The present study aimed to be investigate the comparative genomics and evolutionary divergence of TLR across five vertebrate groups- humans, birds, reptiles, frogs, and fish. The findings provide insights into how TLRs have evolved under selective pressure, shaped by species-specific environments and pathogen exposure. The result from multiple sequence alignments and phylogenetic analysis demonstrated both conservation and divergence in TLR sequence. The intracellular TIR domain showed strong conservation across species, reaffirming a critical role in downstream signaling. However, the extracellular leucine-Rich Repeat (LRR) regions exhibited substantial variation, particularly in non mammalian species. This variation is consistent with the idea that TLRs adapt their pathogen recognition domains in response to environmental microbial diversity. For instance, amphibians and fish displayed lineage-specific expansions of certain TLRs, such as TLR 21 and TLR 22,which are absent in mammals but play essential roles in recognizing nucleic acid ligands in aquatic organisms.

Phylogenetic tree indicated clear clustering of TLRs by type rather than species, supporting the concept of orthology among core TLR genes. However, species-specific paralogs suggest episodes of gene duplication and functional divergence. These results align with previous research indicating that TLRs undergo birth-and-death evolution, where certain genes are maintained under purifying selection, while others evolve under positive selection pressure. The physicochemical analysis also showed variation in parameters like isoelectric point, hydropathicity, and stability, further suggesting functional adaptations of TLR proteins in different species.

The molecular weights of Homo sapiens Such differences may reflect species-specific adaptations, such as variations in extracellular domain length or glycosylation patterns, which could influence ligand recognition and receptor activation (Werlling et al., 2009). Despite these differences, the overall range of molecular weights suggests a conserved structural framework necessary for the maintenance of essential TLR functions across vertebrates.

The theoretical PI property may be crucial for their interaction with ligands and membrane components. Minor shifts in PI, especially in amphibian and reptilian species, might be adaptations to different tissue environments or pathogen pressures (Roach et al.,2005).

The Instability index values were mostly indicative of structural stability higher values indicate a higher tendency for degradation or denaturation. Stable proteins are more likely to retain their structure and function in therapeutic conditions. Here TLRs protein show instability index > 40 it suggests they might be inherently unstable in vitro, Conversely stable TLRs (index <40 might indicate structurally conserved domains essential for Ligand binding and downstream signaling.

Multiple Sequence Alignment :

The multiple sequence alignment of TLR protein sequences from human, frog, bird, fish, and reptile revealed both conserved and variable regions. The conserved residues were mainly located in the intracellular Toll/Interleukin-1 receptor (TIR) domain, highlighting the evolutionary conservation of signaling functionality. In contrast, significant variability was observed in the extracellular leucine-rich repeat (LRR) domains, particularly in surface-exposed regions. These variations likely reflect adaptive changes for pathogen recognition across different species.

Structural visualization confirmed that while the general architecture of TLRs remains conserved, conformational changes in loop regions and ligand- binding sites reflect functional divergence. These changes could potentially influence Receptor-ligand binding affinity, downstream signaling efficiency and immune response specificity in different vertebrates. Importantly, the study underscores the role of ecological and evolutionary pressure n shaping the TLR repertoire of each species. Aquatic vertebrates, such as fish and amphibians, tend to retain a wider diversity of TLR genes likely due to exposure to diverse waterborne pathogens, while birds and reptiles show more specialized adaptations.

Principal Component Analysis performed using PAST software showed clear separation between TLRs from aquatic (fish) and terrestrial species (amphibians, reptiles, birds, and mammals). The first two principal components (PC1 and PC2) explained 71.3% of the total variance in the dataset. PC1 was largely influenced by molecular weight and GRAVY, while PC2 was affected by the instability index and pI. The PCA biplot (Figure 3.1) demonstrated clustering of mammalian and bird TLRs, suggesting similar physicochemical profiles, while fish TLRs occupied a distinct position, indicating functional divergence

This result collaborates earlier phylogenetic studies indicating that fish TLRs have evolved under distinct selective pressures compared to terrestrial vertebrates (Quiniou et al., 2013).

In summary, this study demonstrates that TLRs serve as excellent molecular models to understand immune evolution. The comparative genomic approach revealed how ancient immune components are conserved across vertebrate yet diversified in function to meet species-specific needs. These finding lay the groundwork for future studies exploring host-pathogen interactions, vaccine design, and functional genomics of innate immunity.

CONCLUSION

Toll-like receptors (TLRs) are crucial components of innate immunity, recognizing pathogen-associated molecular patterns and triggering immune responses. This study explored the evolutionary patterns of TLRs across humans, birds, amphibians, reptiles, and Teleost fishes, focusing on how selective pressures shaped their divergence. Comparative genomics, sequence alignments, Phylogenetic analysis, selection pressure detection, and structural modeling revealed that the intracellular TIR domain is highly conserved due to strong purifying selection, while the extracellular LRR domains are more variable, reflecting adaptations to diverse pathogens. Phylogenetic analyses grouped TLRs into functional subfamilies and showed species-specific expansions, especially in teleosts due to genome duplication. Positive selection was mainly found in the extracellular regions, especially at ligand-binding sites and dimerization interfaces. Structural modeling illustrated that even minor amino acid changes can impact ligand recognition. Overall, TLR evolution reflects a balance between maintaining essential signaling functions and adapting to pathogen diversity. These findings deepen our understanding of evolutionary immunology and have implications for vaccine development and immunotherapy. Future research should include more species and functional studies to link mutations to immune function more clearly.

Challenges:

- Limited availability of High quality ancient protein for evolutionary studies.
- Functional validation of variants requires experimental resources.
- Complex interactions between TLR evolution and environmental factors.

APENDIX

1.

sequence allignment TREE.mtsx

2.

sequence allignment TREE.png

3.

RAW SEQ (1).txt

ANALYSIS (1).xlsx

5.

4.

aligned_seq (3).fasta

6. 2a0z (2).pdb

REFERENCE

1. Anderson kv, jürgens g, nüsslein-volhard c. Establishment of dorsal-ventral polarity in the drosophila embryo: genetic studies on the role of the toll gene product. Cell. 1985 oct 1;42(3):779-89.
2. Areal, h., abrantes, j., & esteves, p. J. (2011). Signatures of positive selection in toll-like receptor (tlr) genes in mammals. Bmc evolutionary biology, 11(1), 368.
3. Bagheri, m., & zahmatkesh, a. (2018). Evolution and species-specific conservation of toll-like receptors in terrestrial vertebrates. International reviews of immunology, 37(5), 217–228.
4. Bailey, c. C., zhong, g., huang, i. C., & farzan, m. (2014). Ifitm-family proteins: the cell's first line of antiviral defense. *Annual review of virology*, **1**, 261–283.
   Https://doi.org/10.1146/annurev-virology-031413-085537
5. Bell, j. K., mullen, g. E., leifer, c. A., mazzoni, a., davies, d. R., & segal, d. M. (2003). Leucine-rich repeats and pathogen recognition in toll-like receptors. *Trends in immunology*, 24(10), 528–533.
6. Blaise, b. J., kim, j., & chakraborty, r. (2013). Evolutionary history of toll-like receptor genes in reptiles. *Genome biology and evolution*, 5(9), 1787–1799.
7. Botos, i., segal, d. M., & davies, d. R. (2011). The structural biology of toll-like receptors. Structure, 19(4), 447–459.
8. Brennan, j. J., messerschmidt, j. L., williams, l. M., matthews, b. J., reynoso, m., & gilmore, t. D. (2017).
9. Sea anemone model has a single toll-like receptor that can function in pathogen detection, nf-κb signal transduction, and development. *Proceedings of the national academy of sciences*, 114(47), e10122–e10131.
   Https://doi.org/10.1073/pnas.1711530114

10. Brownlie, r., & allan, b. (2011). Avian toll-like receptors. *Cell and tissue research*, 343(1), 121–130. Https://doi.org/10.1007/s00441-010-1026-0

11. Dheilly, n. M., adema, c. M., raftos, d. A., gourbal, b., grunau, c., & du pasquier, l. (2014). No more non-model species: the promise of next generation sequencing for comparative immunology. *Developmental & comparative immunology*, **45(1)**, 56–66. Https://doi.org/10.1016/j.dci.2014.01.006

12. Grueber, c. E., wallis, g. P., & jamieson, i. G. (2014). Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes. *Plos one*, 9(3), e89632.

13. Guangshuai liu, huanxin zhang, chao zhao, honghai zhang, evolutionary history of the toll-like receptor gene family across vertebrates, genome biology and evolution, volume 12, issue 1, january 2020, pages 3615–3634,

14. Hana velová, maria w gutowska-ding, david w burt, michal vinkler, toll-like receptor evolution in birds: gene duplication, pseudogenization, and diversifying selection, molecular biology and evolution, volume 35, issue 9, september 2018, pages 2170–2184,

15. Huang, s., yuan, s., guo, l., yu, y., li, j., wu, t., ... & yu, y. (2008). Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innovation and complexity. *Nature immunology*, 9(8), 819–825.

16. Imler, j. L., & hoffmann, j. A. (2002). Toll receptors in innate immunity. *Trends in cell biology*, **12(3)**, 134–139. Https://doi.org/10.1016/s0962-8924(01)02298-6

17. Ishengoma, e., agaba, m. Evolution of toll-like receptors in the context of terrestrial ungulates and cetaceans diversification. Bmc evol biol 17, 54 (2017).

18. J.c. roach, g. Glusman, l. Rowen, a. Kaur, m.k. purcell, k.d. smith, l.e. hood, & a. Aderem, the evolution of vertebrate toll-like receptors, proc. Natl. Acad. Sci. U.s.a. 102 (27) 9577-9582, (2005).
19. J.k. bell, i. Botos, p.r. hall, j. Askins, j. Shiloach, d.m. segal, & d.r. davies, the molecular structure of the toll-like receptor 3 ligand-binding domain, proc. Natl. Acad. Sci. U.s.a. 102 (31) 10976-10980, https://doi.org/10.1073/pnas.0505077102 (2005).
20. Janeway, c. A., & medzhitov, r. (2002). Innate immune recognition. *Annual review of immunology*, **20**, 197–216. Https://doi.org/10.1146/annurev.immunol.20.083001.084359
21. Jann, o.c., king, a., corrales, n.l. et al. Comparative genomics of toll-like receptor signalling in five species. Bmc genomics 10, 216 (2009).
22. Jin, m. S., et al. (2007). Crystal structure of the tlr1-tlr2 heterodimer induced by binding of a tri-acylated lipopeptide. Cell, 130(6), 1071–1082.
23. Jumper, j., evans, r., pritzel, a., et al. (2021). Highly accurate protein structure prediction with alphafold. *Nature*, 596, 583–589.
24. Jungi, t.w., farhat, k., burgener, i.a. et al. Toll-like receptors in domestic animals. Cell tissue res 343, 107–120 (2011). Https://doi.org/10.1007/s00441-010-1047-8
25. Kawai, t., akira, s. The role of pattern-recognition receptors in innate immunity: update on toll-like receptors. Nat immunol 11, 373–384 (2010).
26. **Kopp, e., & medzhitov, r. (1999).** The toll-receptor family and control of innate immunity. *Current opinion in immunology*, **11(1)**, 13–18. Https://doi.org/10.1016/s0952-7915(99)80004-7
27. Kumar, h., kawai, t., & akira, s. (2011). Pathogen recognition in the innate immune response. *Biochemical journal*, **420(1)**, 1–16. Https://doi.org/10.1042/bj20111241
28. Lemaitre, b., nicolas, e., michaut, l., reichhart, j. M., & hoffmann, j. A. (1996). The dorsoventral regulatory gene cassette spätzle/toll/cactus controls the potent antifungal response in drosophila adults. *Cell*, 86(6), 973–983.
29. Leulier, f., lemaitre, b. Toll-like receptors — taking an evolutionary approach. Nat rev genet 9, 165–178 (2008). Https://doi.org/10.1038/nrg2303

30. Lu, y.-c., yeh, w.-c., & ohashi, p. S. (2008**).** Lps/tlr4 signal transduction pathway. *Cytokine*, **42(2)**, 145–151. Https://doi.org/10.1016/j.cyto.2008.01.006
31. Matsushima, n., tanaka, t., enkhbayar, p. Et al. Comparative sequence analysis of leucine-rich repeats (lrrs) within vertebrate toll-like receptors. Bmc genomics 8, 124 (2007).
32. Medzhitov, r., et al. (1997). A human homologue of the drosophila toll protein signals activation of adaptive immunity. Nature, 388(6640), 394–397.
33. Mukherjee, s., & karmakar, s. (2022). Toll-like receptor pathways and their crosstalk with other innate immune pathways: evolutionary perspective. *Immunology letters*, 244, 15–24. Https://doi.org/10.1016/j.imlet.2022.01.003
34. O'neill, l. A., golenbock, d., & bowie, a. G. (2013). The history of toll-like receptors — redefining innate immunity. *Nature reviews immunology*, **13(6)**, 453–460. Https://doi.org/10.1038/nri3446
35. Palti, y. (2011**).**toll-like receptors in bony fish: from genomics to function. *Developmental & comparative immunology*, **35(12)**, 1263–1272. Https://doi.org/10.1016/j.dci.2011.03.006
36. Pasare, c., medzhitov, r. (2005). Toll-like receptors: linking innate and adaptive immunity. In: gupta, s., paul, w.e., steinman, r. (eds) mechanisms of lymphocyte activation and immune regulation x. Advances in experimental medicine and biology, vol 560. Springer, boston, ma.
37. Poltorak, a., he, x., smirnova, i., liu, m. Y., van huffel, c., du, x., ... & beutler, b. (1998). Defective lps signaling in c3h/hej and c57bl/10sccr mice: mutations in tlr4 gene. *Science*, 282(5396), 2085–2088.
38. Qi, h., hu, z., zhang, h., et al. (2018). Evolution and expression of the tlr gene family in fish. *Immunogenetics*, 70(7), 429–441.
39. Qi, h., sun, y., sun, x., & xu, h. (2010). Comparative genomics of the toll-like receptor gene family in vertebrates. *Developmental & comparative immunology*, **34(7)**, 763–769. Https://doi.org/10.1016/j.dci.2010.02.008
40. Quiniou, s. M. A., et al. (2013). Evolution of toll-like receptors in vertebrates: fish-specific expansions and adaptive evolution. Bmc evolutionary biology, 13, 194.

41. Quiniou, s.m.a., boudinot, p. & bengtén, e. Comprehensive survey and genomic characterization of toll-like receptors (tlrs) in channel catfish, ictalurus punctatus: identification of novel fish tlrs. Immunogenetics 65, 511–530 (2013).
42. Rebl, a., goldammer, t., & seyfert, h. M. (2010). Toll-like receptor signaling in bony fish. *Veterinary immunology and immunopathology*, 134(3–4), 139–150. Https://doi.org/10.1016/j.vetimm.2009.10.020
43. Roach, j. C., et al. (2005). The evolution of vertebrate toll-like receptors. Pnas, 102(27), 9577–9582.
44. Stejskalova, k., janova, e., splichalova, p. Et al. Twelve toll-like receptor (tlr) genes in the family equidae – comparative genomics, selection and evolution. Vet res commun 48, 725–741 (2024).
45. Takeuchi, o., & akira, s. (2010). Pattern recognition receptors and inflammation. *Cell*, 140(6), 805–820.
46. Temperley, n. D., berlin, s., paton, i. R., griffin, d. K., & burt, d. W. (2008). Evolution of the chicken toll-like receptor gene family: a story of gene gain and gene loss. *Bmc genomics*, 9(1), 62.
47. Uematsu, s., akira, s. Toll-like receptors and innate immunity. J mol med 84, 712–725 (2006).
48. Wang, j., zhang, z., liu, j. Et al. Ectodomain architecture affects sequence and functional evolution of vertebrate toll-like receptors. Sci rep 6, 26705 (2016).
49. Wlasiuk, g., & nachman, m. W. (2010). Adaptation and constraint at toll-like receptors in primates. *Molecular biology and evolution*, **27(9)**, 2172–2186. Https://doi.org/10.1093/molbev/msq103
50. Yeh, d. W., liu, y. L., lo, y. C., yuh, c. H., yu, g. Y., lo, j. F., & chen, n. J. (2013). Toll-like receptor 10 is involved in induction of apoptosis in certain cancer cells. *Journal of biological chemistry*, **288(50)**, 36294–36303. Https://doi.org/10.1074/jbc.m113.504308
51. Yokoyama, w. M. (2005). Missing-self recognition and self-tolerance of natural killer (nk) cells: two sides of the same coin. *Nature reviews immunology*, **5(7)**, 543–551. Https://doi.org/10.1038/nri1633

52. Zhang, s., hu, y., guo, y., guo, x., & zheng, w. (2013). Molecular evolution of tlr1 subfamily in vertebrates: origin and diversification. *Plos one*, **8(4)**, e62350.
Https://doi.org/10.1371/journal.pone.0062350

mbhatta mbhattacharya

Meghana Dalei

Manojit
Zoolgy
Fakir Mohan University

Document Details

Submission ID
trn:oid:::1:3252667111

Submission Date
May 17, 2025, 5:23 PM GMT+5:30

Download Date
May 17, 2025, 5:27 PM GMT+5:30

File Name
msc_thesis_without_anything.docx

File Size
72.9 KB

23 Pages

5,920 Words

36,820 Characters

8% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

30 Not Cited or Quoted 7%
Matches with neither in-text citation nor quotation marks

5 Missing Quotations 1%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

7% Internet sources
5% Publications
1% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.